UNIVERSITE DE NANCY I

****

Numéro d'Enregistrement au C. N. R. S. : A. O. 7569

****

THESE

présentée à l'Université de Nancy I

pour obtenir le grade de

Docteur-Ingenieur

par

ROBERT HORNUNG

Ingénieur I. S. I. N.

****

DISPOSITIF DE RECONNAISSANCE DE LA VOIX PARLEE.

****

soutenue le 14 Février 1973, devant la Commission d'Examen :

Messieurs	R. MAINARD	Président
	J. S. LIENARD	Examinateurs
	R. HUSSON	
	A. FRUHLING	
	M. LAMOTTE	Rapporteur

****

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41555415_0006

DISPOSITIF DE RECONNAISSANCE

DE LA VOIX PARLEE

UNIV.
LEIDEN
BIBL.

AVANT - PROPOS

Ce travail a été réalisé au Laboratoire d'Electricité et d'Automatique de l'Université de Nancy I.

Nous exprimons toute notre gratitude à Monsieur le Professeur A. FRUHLING pour la confiance qu'il nous a témoignée en nous accueillant dans son Laboratoire et pour l'intérêt qu'il a porté à ce travail.

Nous adressons nos vifs remerciements à Monsieur le Professeur R. MAINARD de nous avoir fait l'honneur de présider le Jury, ainsi qu'à Monsieur le Professeur R. HUSSON et à Monsieur J. S. LIENARD, qui a aimablement accepté de se déplacer depuis Paris, d'avoir bien voulu en faire partie.

Nous prions plus particulièrement Monsieur M. LAMOTTE de trouver ici l'expression de notre profonde reconnaissance pour l'aide efficace qu'il n'a cessée de nous apporter dans la conduite de cette réalisation.

Nous savons gré à Madame J. FRUHLING, Mademoiselle J. BOGUET et Monsieur G. VIGNERON de s'être acquittés avec amabilité et diligence de la présentation de ce travail.

Nos remerciements vont également à tous nos camarades du Laboratoire pour les services qu'ils nous ont rendus et pour l'ambiance amicale qu'ils y font régner.

* * * *

TABLE DES MATIERES

	Page
INTRODUCTION.	1
CHAPITRE I : DESCRIPTION DU PROCESSUS DE RECONNAISSANCE.	3
I.1. Prétraitement	3
I.2. Système de reconnaissance	5
I.2.1. Système de reconnaissance par apprentissage	5
I.2.2. Entrées de la matrice d'apprentissage	7
I.2.3. Matrice d'apprentissage	8
I.2.3.1. Discrimination linéaire	9
I.2.3.2. Application à la matrice	9
I.2.4. Identification de maximum	11
CHAPITRE II : REALISATION DU SYSTEME DE RECONNAISSANCE.	12
II.1. Fonctionnement général	12
II.2. Entrées de la matrice d'apprentissage	16
II.2.1. Calibrage des signaux d'entrée des intégrations	17
II.2.1.1. Calibrage C1	18
II.2.1.2. Calibrage C2	18
II.2.1.3. Calibrage C3	19
II.2.1.4. Réalisation à l'aide de circuits intégrés	21
II.2.2. Elaboration des entrées E_j^+ et E_j^- de la matrice d'apprentissage	25
II.2.3. Calcul des paramètrss intervenant dans l'élaboration des entrées de la matrice d'apprentissage	30
II.2.3.1. Position du problème	30
II.2.3.2. Calcul de la constante de temps RC d'un intégrateur	31
II.2.3.3. Calcul de la largeur θ_f de l'impulsion S_f	33
II.2.3.4. Calcul de la constante de temps des monostables 1, 2, 3, 4 et 5	33
II.3. Calcul des résistances de la matrice d'apprentissage	35
II.4. Identification de maximum	39
II.4.1. Fonctionnement général et synchronisation d'un cycle de tests	40
II.4.2. Fonctionnement du multiplexeur	42
II.4.3. Convertisseur analogique-numérique	44
II.4.4. Comparateur digital	46
II.4.4.1. Principe de la comparaison	46
II.4.4.2. Réalisation du comparateur	47

CHAPITRE III : REALISATION D'UN SYSTEME DE SEGMENTATION PHONEMIQUE DE LA PAROLE. 48
III.1. Principe de la segmentation 49
III.2. Réalisation du système de segmentation 53
III.2.1. Fonctionnement général 53
III.2.2. Synchronisme des opérations 56
III.2.3. Circuits intégrés utilisés 57

CHAPITRE IV : EXPLOITATION DU SYSTEME DE RECONNAISSANCE ET DE SEGMENTATION - RESULTATS. 59
IV.1. Saisie des informations sur calculateur T 2000 59
IV.2. Résultats fournis par l'ensemble de segmentation 62
IV.3. Vérification du bon fonctionnement de l'ensemble de traitement - Résultats en reconnaissance 63

CONCLUSION.

BIBLIOGRAPHIE.

INTRODUCTION

La parole est un phénomène complexe, difficile à analyser. Les recherches dans ce domaine, pendant longtemps limitées à la synthèse et à la transmission de la voix parlée, se sont étendues, depuis une vingtaine d'années, à la reconnaissance automatique de celle-ci.

Le but, généralement poursuivi en reconnaissance, est la réalisation d'une machine à reconnaître la parole en temps réel, c'est-à-dire la réalisation d'une machine capable de comprendre un mot ou une phrase et de les transformer par exemple en une opération (ordre de travail) ou en un message écrit et visualisé sur les terminaux usuels (machine à écrire, imprimante, console de visualisation). Dans la plupart des systèmes réalisés, la reconnaissance d'un message parlé s'établit au niveau de formes élémentaires telles que la syllabe ou le mot.

Nous décrivons dans ce travail la réalisation d'un appareil de reconnaissance où le traitement de la parole s'effectue en temps réel, au niveau du phonème, plus petite unité phonologique que l'on puisse isoler dans la parole. Cette réalisation concrétise un processus de reconnaissance simulé sur ordinateur CII 10070 mettant en oeuvre une matrice d'apprentissage. Les données de base du traitement sont fournies par un dispositif de prétraitement déjà réalisé qui codifie le signal vocal sous une forme numérique.

Dans une première partie, nous rappelons brièvement comment s'opère le prétraitement numérique du signal de parole, puis nous expliquons quelle est la configuration du processus de reconnaissance simulé sur ordinateur. Trois fonctions y seront distinguées : l'élaboration d'une forme intermédiaire à partir des données numériques issues

du prétraitement, la discrimination de cette forme par une matrice d'apprentissage et l'identification sur les sorties de la matrice, du phonème reconnu.

La deuxième partie est consacrée à la réalisation proprement dite du système de reconnaissance. Elle s'attache à la description de la transcription technologique des trois fonctions précitées.

La troisième partie est réservée à la présentation d'un dispositif de segmentation phonémique d'un message parlé, basée sur le repérage de la transition entre un régime de parole soutenu et un régime de parole transitoire.

Enfin, dans la dernière partie, nous soulignons comment peuvent être vérifiés, à l'aide d'un calculateur industriel T 2000 , les résultats obtenus et nous les commentons.

* * *
*

CHAPITRE I - DESCRIPTION DU PROCESSUS DE RECONNAISSANCE.

La reconnaissance de la voix parlée consiste à affecter une classe d'appartenance, ou catégorie, à une forme vocale présentée pouvant être le phonème, le mot ou la phrase.

Dans cette étude, la forme vocale élémentaire considérée est le phonème, plus petite unité phonologique que l'on puisse isoler dans la parole. Le nombre des principaux phonèmes en français étant environ de 30, on définit donc 30 catégories, la reconnaissance établissant l'appartenance des phonèmes constitutifs du signal vocal à la catégorie qui leur est impartie [1 , 2].

Il y a lieu de distinguer deux phases dans le processus qui est mis en œuvre ici :

- le prétraitement du signal de parole : motivé par la nécessité de quantifier ce signal pour en éliminer les informations superflues et en extraire les caractéristiques les plus descriptives ; ce prétraitement est réalisé par analyse fréquentielle sur un banc de filtres.

- le traitement proprement dit utilisant les données issues du prétraitement et les appliquant à un système de reconnaissance par apprentissage.

Nous examinerons successivement, dans ce chapitre, ces deux centres d'intérêt.

I.1. Prétraitement.

L'étude et la réalisation du système de prétraitement ont fait l'objet de travaux antérieurs [3 , 4] ; aussi, c'est brièvement que nous rappellerons le fonctionnement de ce dispositif.

L'aspect global d'un phonème est défini dans l'espace à trois dimensions amplitude - temps - fréquence, par les formants, zones de fréquence où l'amplitude est maximale et par leurs déplacements en fréquence au cours du temps (planche 1).

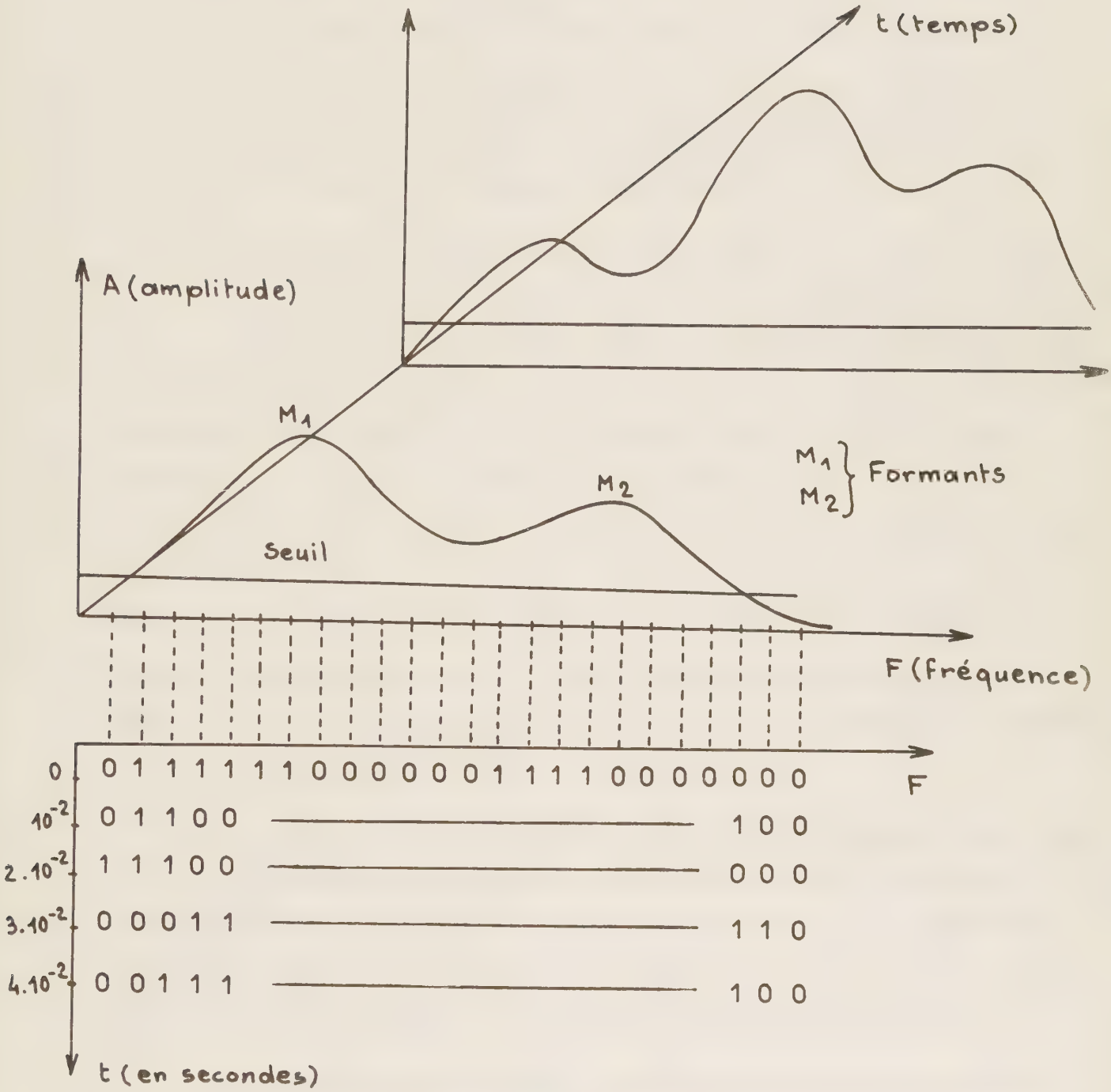

t (temps)
A (amplitude)
M1
M2
Formants
Seuil
F (fréquence)
0 0 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0
F
10^-2 0 1 1 0 0 1 0 0
2.10^-2 1 1 1 0 0 0 0 0
3.10^-2 0 0 0 1 1 1 1 0
4.10^-2 0 0 1 1 1 1 0 0
t (en secondes)

Planche 1

L'analyseur réalisé quantifie le signal vocal dans ce système d'axes ; son schéma synoptique est donné planche 2.

- La fréquence d'échantillonnage dans le temps est de 100 Hz. Elle est suffisante pour suivre l'évolution des transitoires de la parole dont la durée est généralement comprise entre 25 et 50 ms.

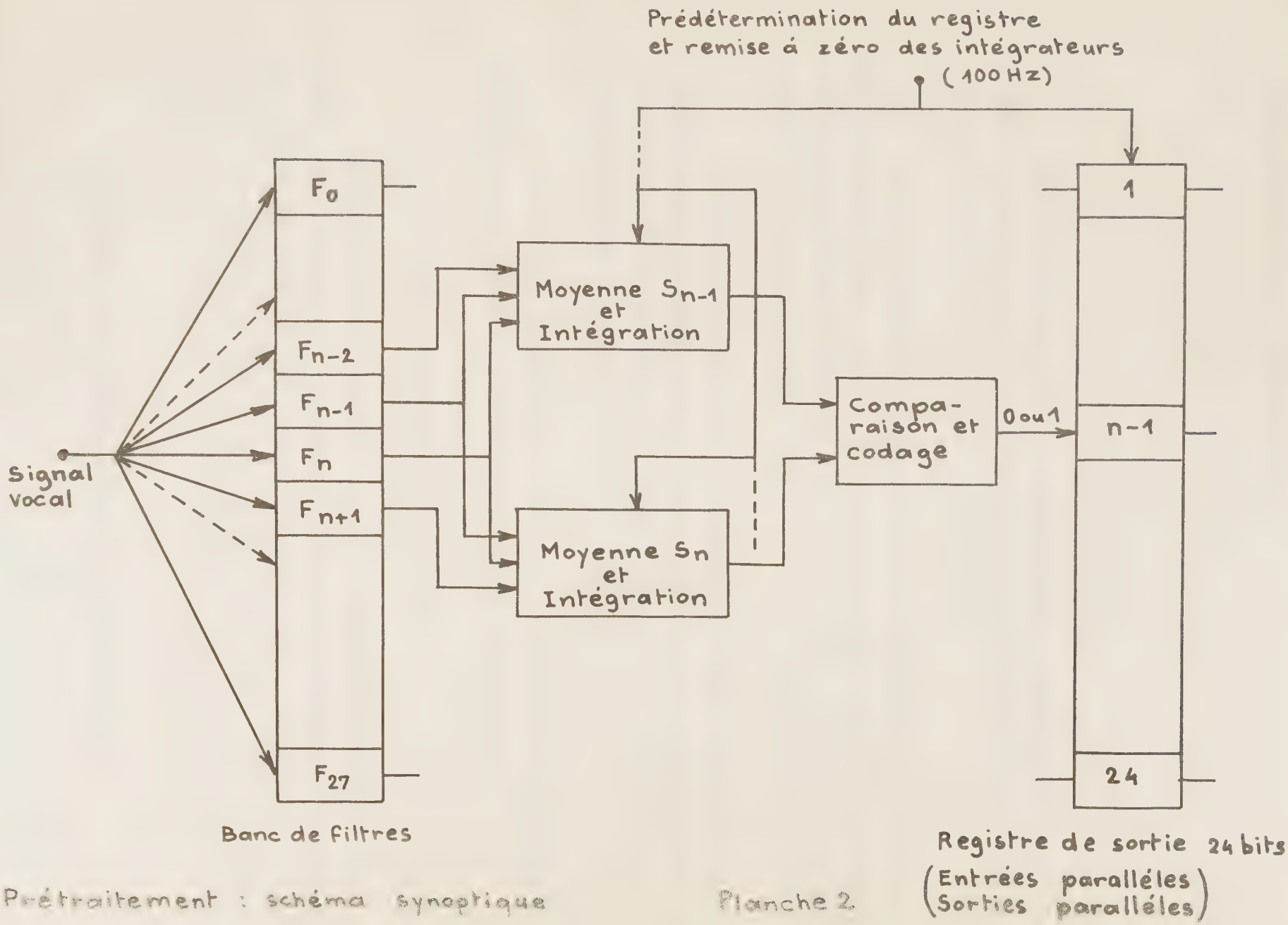
- L'analyse spectrale (échantillonnage en fréquence) est effectuée par un banc de filtres passe - bande disposés en parallèle et dont les fréquences de résonance s'échelonnent de 100 à 7000 Hz.

- Les variations d'amplitude sont discrétisées par comparaison des tensions de sortie redressées V_{n-1} et V_n de deux filtres consécutifs F_{n-1} et F_n. Chaque cellule de comparaison élabore l'information binaire codée 1 dans le cas de la croissance $(V_n > V_{n-1})$ et 0 dans le cas de la décroissance $(V_n < V_{n-1})$ ou encore dans le cas où les deux tensions V_n et V_{n-1} sont inférieures à un seuil choisi.

En réalité, les signaux comparés ne sont pas V_{n-1} et V_n, mais respectivement $S_{n-1} = \frac{V_{n-2}}{2} + V_{n-1} + \frac{V_n}{2}$ et $S_n = \frac{V_{n-1}}{2} + V_n + \frac{V_{n+1}}{2}$, sommes pondérées des tensions de sortie redressées de trois filtres consécutifs. On élimine ainsi les variations fines de la courbe amplitude - fréquence pour faire apparaître les formants avec plus de netteté.

Ces deux signaux sont ensuite intégrés durant le pas d'échantillonnage (10 ms) afin de diminuer l'effet de la composante alternative subsistant à la sortie des filtres. Le premier intégrateur qui atteint un seuil prédéterminé définit le filtre dont la sortie est la plus élevée. Le résultat est alors codifié d'après les critères définis ci-dessus et stocké dans le registre de sortie.

Nous avons préféré traduire ce codage par les variations relatives de l'amplitude, car les caractères de la parole dépendent beaucoup plus de la position des formants que de leur valeur absolue, au demeurant variable selon l'intensité du signal.



Prétraitement : schéma synoptique

Planche 2

Ce prétraitement donne du signal vocal continu une forme discrète réduite à 2400 bauds, selon les transformations :

- temps : fréquence d'échantillonnage 100 Hz
- amplitude : deux niveaux (0 ou 1)
- fréquence : 24 comparaisons spectrales.

Les 24 informations binaires disponibles toutes les 10^{-2} s aux sorties du registre constituent les données de base du traitement proprement dit, que ce dernier soit simulé ou en temps réel.

Pour la simulation, ces données sont enregistrées, par groupe de six, sur ruban perforé. Quatre enregistrements successifs sont ainsi nécessaires pour analyser une même séquence de parole, le regroupement de ces quatre enregistrements partiels se faisant par programmation.

A l'issue du prétraitement, un phonème présenté au système de reconnaissance s'identifie donc à un tableau rectangulaire rempli de "0" ou de "1" (planche 1). On y dénombre 24 lignes, et autant de colonnes qu'il y a de périodes d'échantillonnage dans la durée du phonème.

La reconnaissance consiste à grouper, dans une même catégorie, des formes présentant une répartition similaire de nuages de "1".

I.2. Système de reconnaissance. [7 , 8]

Dans cette partie, nous décrivons le système de reconnaissance utilisé. Nous motivons, en particulier, la nécessité de réaliser la reconnaissance par apprentissage et, celle d'élaborer une forme intermédiaire à partir des données brutes issues du prétraitement. Enfin, nous explicitons comment la simulation sur ordinateur a permis le calcul des paramètres du système.

I.2.1. Système de reconnaissance par apprentissage.

La structure d'un processus de reconnaissance de formes est, en premier lieu, impliquée par la nature des formes qui lui sont présentées. Deux cas simples sont à envisager [5 , 6] .

a) - Si les formes sont décrites avec précision à l'aide de caractéristiques bien choisies, la reconnaissance établit sans ambiguité l'appartenance de ces formes à leur catégorie, par comparaison avec une forme type de la même catégorie.

b) - Si, par contre, le lien entre les caractéristiques des formes et leur catégorie d'appartenance n'est pas évident, on est obligé, dans un premier temps, de "sensibiliser" le système aux formes qui lui sont présentées. On a alors recours à un système de reconnaissance par apprentissage.

Les formes vocales s'apparentent à celles du deuxième cas. En effet, le même phonème ne présente pas toujours le même aspect : il existe des différences dues aux phonèmes précédents et suivants, dues également aux locuteurs et par conséquent, les critères de reconnaissance ne s'imposent pas toujours de façon systématique.

En général, on se fixe la forme des fonctions de discrimination, donc la structure du processus de reconnaissance par apprentissage. La reconnaissance s'opère ensuite en deux phases : la phase dite d'apprentissage puis la phase leçon.

Durant la première phase, les formes à reconnaître sont déjà identifiées. L'apprentissage consiste à modifier les paramètres du système (et par suite les coefficients des fonctions de discrimination) jusqu'à ce que, pour une succession de formes présentées, les réponses du système soient correctes.

Durant la seconde phase, phase de reconnaissance proprement dite, le système classe des formes non identifiées suivant les critères établis au cours de l'apprentissage.

En ce qui concerne le système de reconnaissance utilisé, nous distinguerons trois parties dont le schéma général est donné planche 3 :

- un ensemble de traitement des informations numériques fournies par le prétraitement, qui élabore les formes intermédiaires nécessaires à l'unité suivante ;

- une unité de reconnaissance proprement dite, constituée par une matrice d'apprentissage ;

- un identificateur de maximum, détectant, sur les sorties de la matrice, les phonèmes reconnus.

I.2.2. Entrées de la matrice d'apprentissage.

Nous avons vu que l'acquisition des données digitales issues du prétraitement se fait pour 24 fréquences tous les centièmes de seconde. En supposant la durée moyenne d'un phonème égale à 10^{-1} s , une forme à reconnaître est donc décrite par environ 240 positions binaires qui pourraient constituer les entrées de la matrice d'apprentissage.

Le nombre de ces entrées est, en fait, considérablement réduit, si l'on tient compte de la continuité du signal vocal dans le temps, un phonème n'étant perçu que s'il dure suffisamment longtemps. Aussi, il semble avantageux de créer, à partir de la forme quantifiée, une forme intermédiaire qui caractérise simultanément l'évolution instantanée et celle, quelque peu antérieure, du signal.

Ce résultat est obtenu en disposant sur chaque entrée, un intégrateur doué de fuites. L'intégrateur se charge d'une quantité constante chaque fois qu'un niveau "1" se présente (son état reste inchangé s'il s'agit d'un niveau "0"), puis perd, entre deux échantillonnages successifs, une charge proportionnelle à celle emmagasinée.

L'introduction des fuites présente, en outre, les avantages suivants :

- il n'est pas nécessaire de remettre à zéro l'intégrateur à la fin de chaque phonème ;

- pour les voyelles, souvent caractérisées par une succession ininterrompue de "1", il s'opère une certaine normalisation en temps : le niveau de saturation de l'intégrateur est atteint rapidement, que le phonème soit long ou bref.

Chacune des 24 sorties du prétraitement est mise en correspondance avec un des intégrateurs I_1 à I_{24} par l'intermédiaire d'un registre tampon (planche 3).

De plus, on ajoute :

- trois intégrateurs avec fuites, I_{25} à I_{27} qui se chargent proportionnellement au nombre de "1" présents respectivement dans les zones des basses, moyennes et hautes fréquences, c'est-à-dire dans les trois octets composant le mot binaire stocké dans le registre tampon ;

- un intégrateur I_{28}, qui se charge quand la somme arithmétique des bits du registre est nulle et qui détecte les silences ;

- un intégrateur I_{29}, qui fonctionne quand la somme des bits est supérieure à 7 et un autre I_{30} lorsque celle-ci est inférieure ou égale à 7, ces deux intégrateurs indiquant respectivement les zones de forte et de faible intensité (séparation voyelles - consonnes).

La matrice possède donc 30 entrées ; les six dernières, définies ci-dessus, expriment la répartition approximative des bits égaux à 1 dans le registre.

La simulation sur ordinateur des opérations introduites jusqu'ici, se réalise simplement par addition puis division, pour les 24 premiers intégrateurs ; un comptage précède ces mêmes opérations, pour les six derniers.

En fait, ce n'est pas la grandeur de sortie d'un intégrateur qui constitue une des entrées de la matrice d'apprentissage, mais l'écart entre cette grandeur et sa valeur moyenne, calculée auparavant, par simulation, sur une séquence de parole suffisamment longue ; cet écart fait ressortir les traits caractéristiques qui servent à la discrimination des formes.

Remarquons, pour conclure cette partie, qu'élaborer les entrées de la matrice par simulation revient à transposer une information binaire en un nombre positif ou négatif.

I.2.3. Matrice d'apprentissage.

Deuxième bloc de la chaîne de traitement, la matrice d'apprentissage met en œuvre des fonctions de discrimination linéaires qui servent à l'identification des formes.

I.2.3.1. Discrimination linéaire [1] .

Le problème de la reconnaissance consiste à classer une forme X à n composantes x_j dans une des N catégories possibles. Cette forme se représente par un point dans l'espace R^n et les fonctions de discrimination linéaires associées à chaque catégorie i ont pour expression :

$$F_i(X) = \sum_{j=1}^{n} P_{ij} \; x_j + P_{i,n+1}$$

Une forme X appartient à la catégorie k si

$$F_k(X) > F_i(X) \quad \text{pour } i = 1, \ldots, N \quad \text{et} \quad i \neq k$$

Cette dernière relation équivaut à partager l'espace R^n en volumes affectés chacun à une catégorie déterminée. La limite de décision entre les catégories k et ℓ se traduit par $F_k(X) = F_\ell(X)$.

Dans le cas de fonctions de discrimination linéaires, cette condition limite s'écrit

$$\sum_{j=1}^{n} (P_{kj} - P_{\ell j}) \, x_j + P_{k,n+1} - P_{\ell,n+1} = 0$$

équation d'un hyperplan dont les pentes suivant les axes sont données par $(P_{kj} - P_{\ell j})$ et le décalage à l'origine par $(P_{k,n+1} - P_{\ell,n+1})$.

Au cours de l'apprentissage, les coefficients P_{ij} sont modifiés selon un algorithme de correction, de façon à déplacer au mieux les hyperplans qui doivent séparer les ensembles de points groupés par catégorie.

I.2.3.2. Application à la matrice.

Les fonctions linéaires de discrimination sont formées dans une matrice qui possède 30 colonnes et 30 lignes. Les colonnes sont les entrées EM_j décrites précédemment, à l'exception de la première entrée qui a été remplacée par un terme constant K ; les lignes sont les sorties SM_i obtenues en faisant la somme pondérée des entrées.

L'expression analytique d'une sortie est donc

$$SM_i = K.P_{1i} + \sum_{j=2}^{30} P_{ij}.EM_j$$

dans laquelle les P_{ij} sont les coefficients de pondération ou poids de la matrice, représentés sur la planche 3, à l'intersection des lignes et des colonnes.

Chaque ligne étant affectée par convention à un phonème, on établit l'appartenance d'une forme présentée à un des phonèmes, quand sur la ligne correspondante, la sortie est plus élevée que la sortie de toutes les autres lignes.

Les coefficients de pondération sont déterminés par apprentissage.

Dans un premier temps, la segmentation, c'est-à-dire le découpage phonémique de la séquence de parole traitée, ainsi que l'identification des phonèmes la constituant, sont effectuées manuellement à partir des formes numériques issues du prétraitement. Cette séquence contient bien évidemment tous les phonèmes à reconnaître.

L'apprentissage se déroule ensuite de la façon suivante :

- un phonème est présenté ;

- si la réponse est correcte, c'est-à-dire si le critère de reconnaissance explicité précédemment est satisfait, on ne modifie pas les poids de la matrice ;

- s'il y a erreur, les poids de la ligne qui aurait dû être détectée sont augmentés d'une valeur proportionnelle aux entrées, ceux de la ligne erronée sont diminués.

Cette opération se répète pour tous les phonèmes de la séquence.

Pour chaque cycle d'apprentissage, on évalue le taux de reconnaissance ; si ce taux est inférieur à un seuil choisi, on recommence un cycle de corrections.

Ainsi, au terme de l'apprentissage par simulation, les poids de la matrice sont définitivement fixés ; nous verrons que dans la réalisation ils sont matérialisés par des résistances.

I. 2. 4. Identificateur de maximum.

Nous avons déjà spécifié l'utilisation de l'identificateur de maximum en tant qu'unité de décision du système. Sa fonction est en effet d'exprimer le phonème reconnu à la fin de chaque échantillonnage. Ce phonème est, d'après le critère de reconnaissance choisi, celui affecté à la ligne dont la sortie est la plus élevée par rapport à celle de toutes les autres lignes.

La simulation de cette opération se réalise par comparaison de toutes les sorties de la matrice et ne présente aucune difficulté.

On peut ajouter pour conclure ce chapitre, que les concepts mis en oeuvre dans le système de reconnaissance ont volontairement été définis de la manière la plus simple possible, puisque notre objectif est de construire un appareillage.

Cette réalisation, décrite dans le chapitre suivant, n'a été possible qu'à partir des résultats fournis par la simulation du processus sur ordinateur. On a pu, grâce à celle-ci, connaître à tous moments l'ordre de grandeur des variables internes du système et, en particulier, mener à bien le calcul de paramètres tels que les valeurs moyennes des sorties des intégrateurs et les poids de la matrice.

. . . . . .

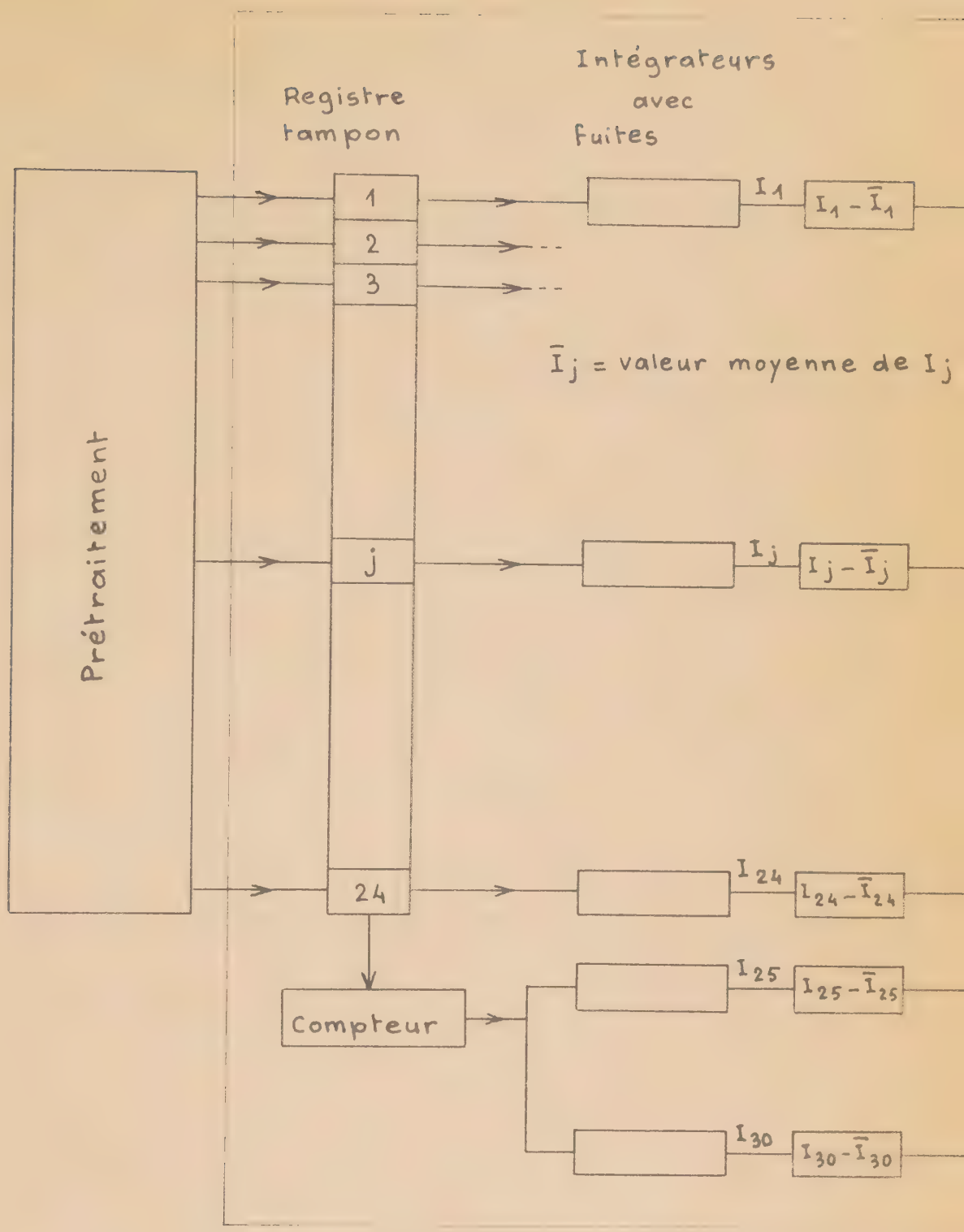
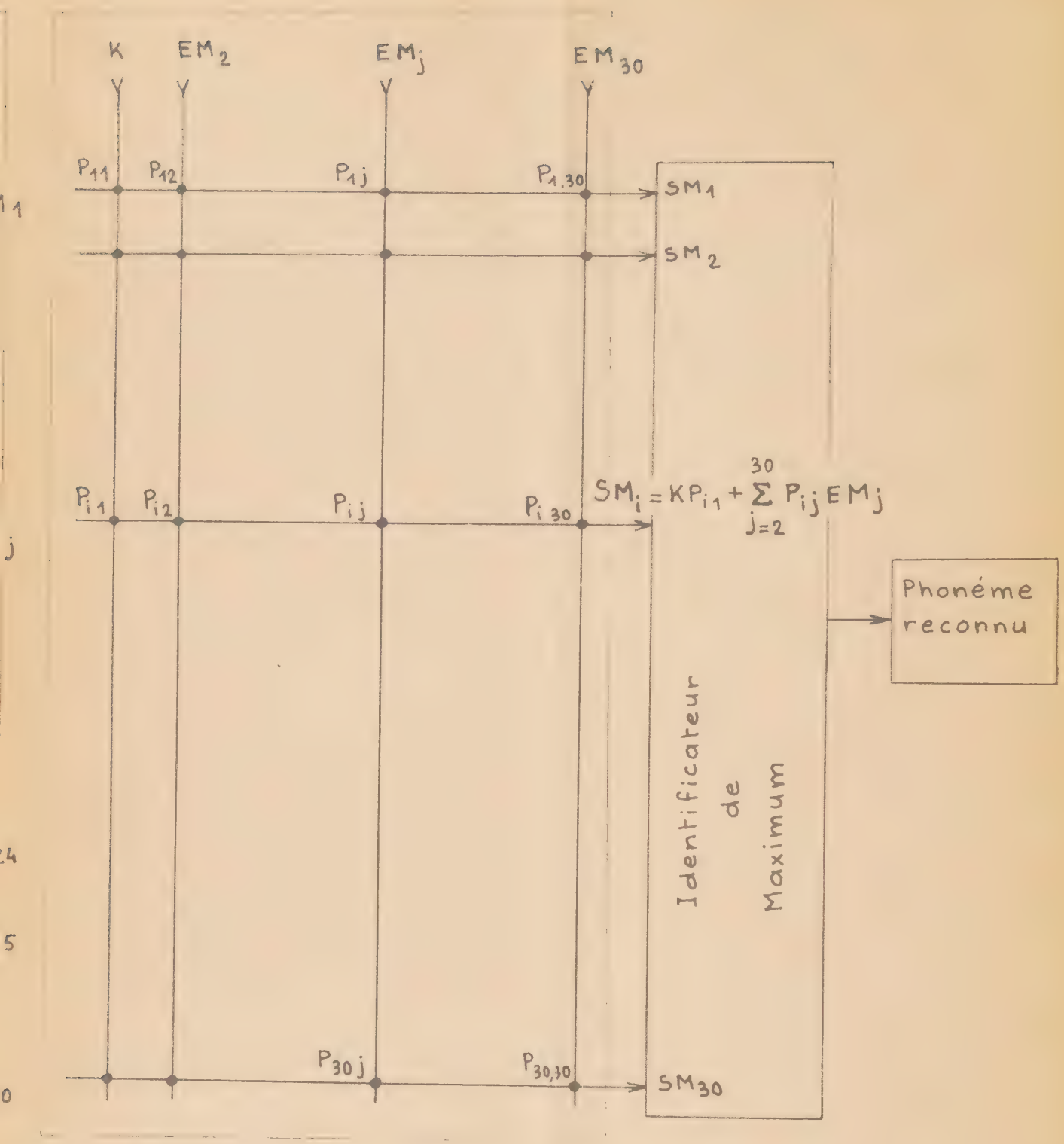


Schéma général du systé



Matrice d'apprentissage

de reconnaissance

Planche 3

CHAPITRE II - REALISATION DU SYSTEME DE RECONNAISSANCE.

Puisque les informations fournies par le prétraitement se présentent sous forme digitale, nous pouvions envisager de construire deux genres de dispositifs : un dispositif de calcul numérique, ou un dispositif de calcul hybride, tantôt numérique, tantôt analogique.

La première solution correspond en fait à la réalisation d'un calculateur muni des quatre opérations élémentaires et d'une mémoire de capacité non négligeable. Nous l'avons naturellement rejetée.

La deuxième offre les avantages du calcul numérique (ordonnancement aisé des opérations, mémorisation des informations), conjugués à ceux du calcul analogique (rapidité dans l'exécution des opérations, affichage simple des paramètres). Elle se prête donc particulièrement bien à notre système, pour lequel le nombre de paramètres introduits est important.

Le processus de reconnaissance réalisé a la même architecture que celui simulé sur ordinateur.

On explicitera successivement dans ce chapitre, son fonctionnement général, puis la technologie de ses diverses parties : le bloc élaborant les entrées de la matrice, la matrice d'apprentissage et l'identificateur de maximum.

II.1. Fonctionnement général.

La succession des opérations de traitement se déroule toutes les 10^{-2}s, au rythme de l'échantillonnage du signal vocal.

On peut aisément la décrire à partir du schéma synoptique donné planche 4 et qui représente la transposition électronique du système simulé sur ordinateur (planche 3).

Au début de chaque cycle de traitement, on dispose sur les sorties du prétraitement, d'un mot de 24 bits représentatif du signal de parole 10 ms plus tôt. Ce mot est transféré en parallèle dans le registre R par application du signal de prédétermination H_{100}.

L'opération d'intégration qui doit suivre et qui numériquement s'opère par addition, se réalise ici analogiquement par intégration d'impulsions positives calibrées en largeur et en amplitude ; ce sont les signaux r_1 , r_2 et r_3 . Cette opération s'effectue en trois étapes successives, chacune relative à un groupe d'intégrateurs :

- les vingt quatre premiers, on le sait, sont affectés aux 24 sorties du registre R . L'état binaire s_j de ces sorties valide (quand $s_j = 1$) ou inhibe (quand $s_j = 0$) l'intégration. Le signal intégré s^*_j est donc le signal de sortie des portes ET de calibrage C1 dans lesquelles sont composés les produits logiques $s_j . r_1$;

- les trois intégrateurs suivants sont chacun relatif à un des 3 octets constituant le mot stocké dans le registre et doivent se recharger proportionnellement au nombre de 1 contenu dans l'octet qui leur est respectif. On provoque alors un décalage global du mot vers le bas à la cadence d'une autre horloge H ; durant les huit premiers décalages, les trois octets sortent en série sur les sorties 9 , 17 et 25 du registre. Le signal d'entrée de chacun des trois intégrateurs correspond ici respectivement à chacun des produits $s_9 . r_2$, $s_{17} . r_2$, $s_{25} . r_2$ formés dans les portes ET de calibrage C2, où l'impulsion calibrée r_2 est répétée 8 fois en synchronisation avec H ;

- au cours du décalage global, on compte également le nombre N de 1 se présentant sur la 25ème sortie du registre, les impulsions d'incrémentation du compteur étant des impulsions d'horloge validées ou non dans la porte I par l'état s_{25} de cette sortie. A l'issue du comptage, on effectue en parallèle trois tests sur N :

$$N = 0$$
$$0 < N \leq 7$$
$$N > 7$$

Un seul de ces tests est positif et c'est l'intégrateur qui lui est affecté qui réalise l'intégration de l'impulsion r_3 appliquée aux portes ET de calibrage C3.

La largeur des impulsions r_1 , r_2 et r_3 est directement proportionnelle à la constante numérique ajoutée pour simuler l'intégration sur les différentes voies.

Les fuites sont ensuite provoquées simultanément sur les 30 intégrateurs par la fermeture des interrupteurs m . Cette fermeture, maintenue durant l'application de l'impulsion calibrée en temps S_f , entraîne la décharge de la capacité C dans la résistance R' d'une quantité d'électricité proportionnelle à la charge initiale de C . Lorsque S_f disparaît, la tension de sortie d'un intégrateur a pour valeur $- V_j$; elle restera constante jusqu'à la fin du cycle.

Les deux amplificateurs opérationnels qui suivent permettent d'élaborer les tensions E_j^+ et E_j^- appliquées à la $j^{ième}$ colonne d'entrée de la matrice d'apprentissage :

- le premier est monté en sommateur inverseur de gain 1 et sa sortie E_j^+ correspond à l'écart entre la tension V_j et sa valeur moyenne $\bar{V}_j$ ($\bar{V}_j$ déterminée par simulation est ici concrétisée par affichage potentiométrique).

- le deuxième, monté en inverseur simple de gain 1 génère le signal E_j^- opposé de E_j^+ .

Ainsi :

$$E_j^+ = (V_j - \bar{V}_j)$$

$$E_j^- = - E_j^+ = - (V_j - \bar{V}_j)$$

La matrice est un réseau de résistances effectuant sur chaque ligne la combinaison linéaire des grandeurs d'entrée. Les résistances qui la composent ont été calculées, on le verra en détail par la suite, en identifiant les équations des sorties de la matrice simulée sur ordinateur,

$$SM_i = K.P_{i1} + \sum_{j=2}^{30} P_{ij}.EM_j$$

à celles de la matrice réalisée, que l'on peut écrire sous la forme

$$S_i = K.f_{i1}(R_{ij}) + \sum_{j=2}^{30} f_{ij}(R_{ij}).E_j$$

En simulation, le signe des poids est positif ou négatif. Dans la réalisation où les fonctions $f_{ij}(R_{ij})$ ne peuvent être que positives, on est conduit à traduire le signe de $f_{ij}(R_{ij})$ par celui des E_j analogiques. Ceci explique le dédoublement de chaque entrée en deux entrées opposées, et le positionnement des résistances schématisé planche 4 ; les résistances de la ligne i par exemple, ont toutes une de leur extrémité connectée à cette ligne, la deuxième étant affectée à une colonne E_j^+ si le poids P_{ij} correspondant est positif et à une colonne E_j^- dans le cas contraire. Les courants qui circulent dans ces différentes résistances créent la tension S_i aux bornes de la résistance de charge R_L.

La dernière opération d'un cycle de traitement est l'identification de la sortie de matrice présentant le potentiel le plus élevé. On procède pour cela à la scrutation successive de toutes les voies d'entrée de l'identificateur, en correspondance directe avec les sorties de la matrice, et on réalise sur chacune d'elles la séquence d'opérations suivante (voir planche 4) :

- la tension analogique S_i générée sur la voie sélectionnée est convertie en tension numérique, puis transférée en parallèle dans le registre I ; le numéro de cette voie est mémorisé dans un compteur ;

- les mots binaires RI et RII contenus respectivement dans le registre I et dans un registre II où se trouve à tout moment la dernière

valeur maximale détectée, sont ensuite comparés ; deux cas sont alors envisagés :

a) - Si RI > RII le comparateur digital valide le transfert parallèle de RI dans le registre II et simultanément celui du contenu du compteur dans un registre III.

b) - Si RI $\leqslant$ RII aucun transfert n'est ordonné.

Ainsi, à la fin du multiplexage des 30 voies, le numéro du phonème reconnu est obtenu par lecture parallèle du registre III.

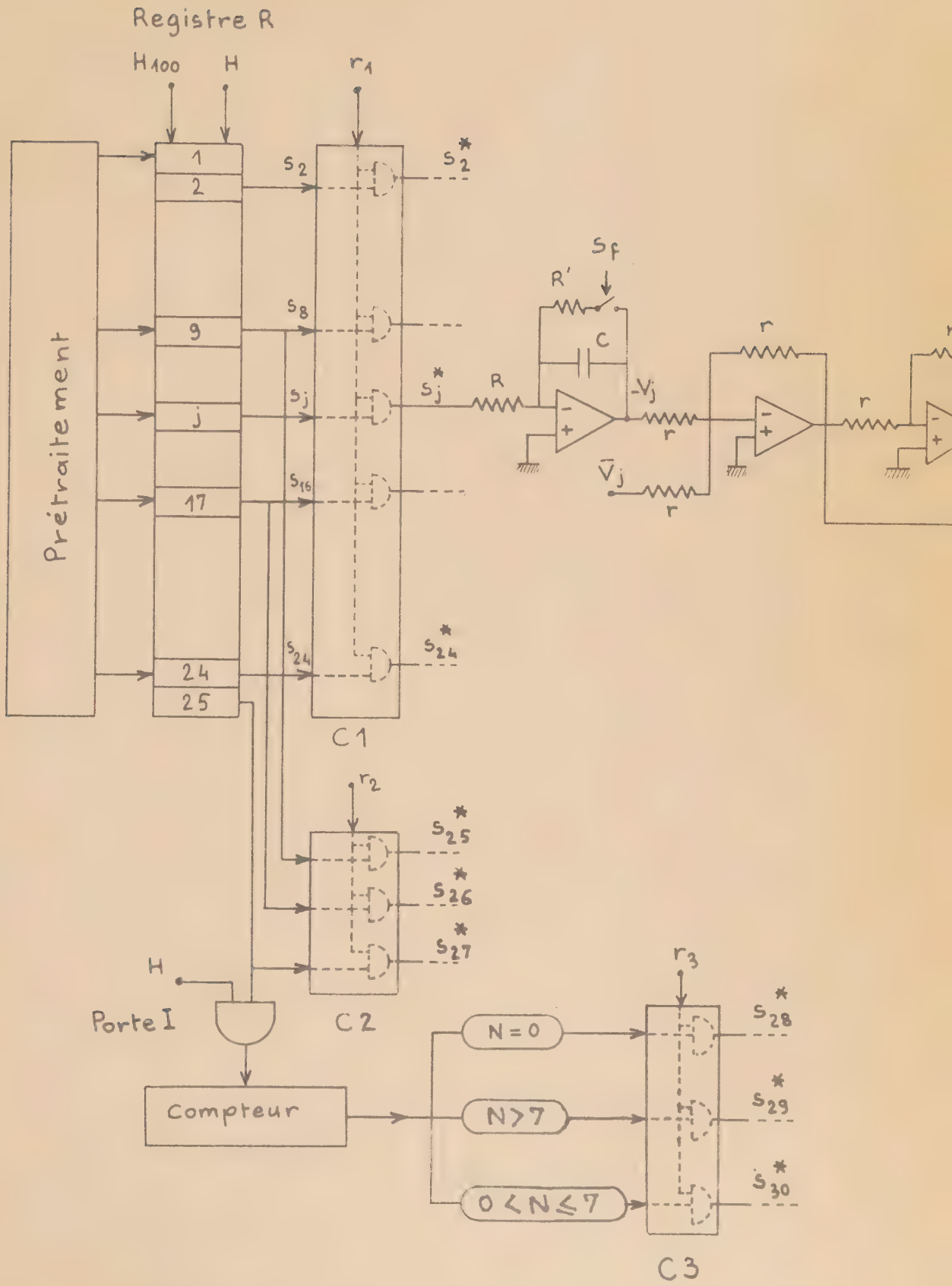
En résumé, l'ordonnancement des opérations se définit comme suit :

- mise en forme des signaux d'entrée des 30 intégrateurs, soit en d'autres termes les calibrages C1 , C2 et C3 ;
- intégration de ces divers signaux au moment où ils sont générés ;
- application des fuites simultanément sur les 30 voies ;
- formation des entrées E_j^+ et E_j^- de la matrice d'apprentissage ;
- pondération linéaire de ces entrées qui se traduit sur chacune des sorties de la matrice par une différence de potentiel S_i ;
- identification de la sortie présentant le potentiel S_i le plus élevé.

Un tel cycle se déroule durant les 10 ms qui séparent deux instants d'échantillonnage.

II.2. Entrées de la matrice d'apprentissage.

Les entrées de la matrice d'apprentissage, on l'a vu, sont des informations analogiques élaborées à partir d'informations numériques. Le dispositif réalisant cette transposition comprend donc deux ensembles :



Registre R
H100
H
r1
Prétraitement
1
2
9
j
17
24
25
s2
s8
sj
s16
s24
s2*
sj*
s24*
C1
Sf
R'
C
R
-Vj
r
V̄j
r
r
r
r2
s25*
s26*
s27*
C2
H
Porte I
Compteur
N=0
N>7
0<N≤7
r3
s28*
s29*
s30*
C3

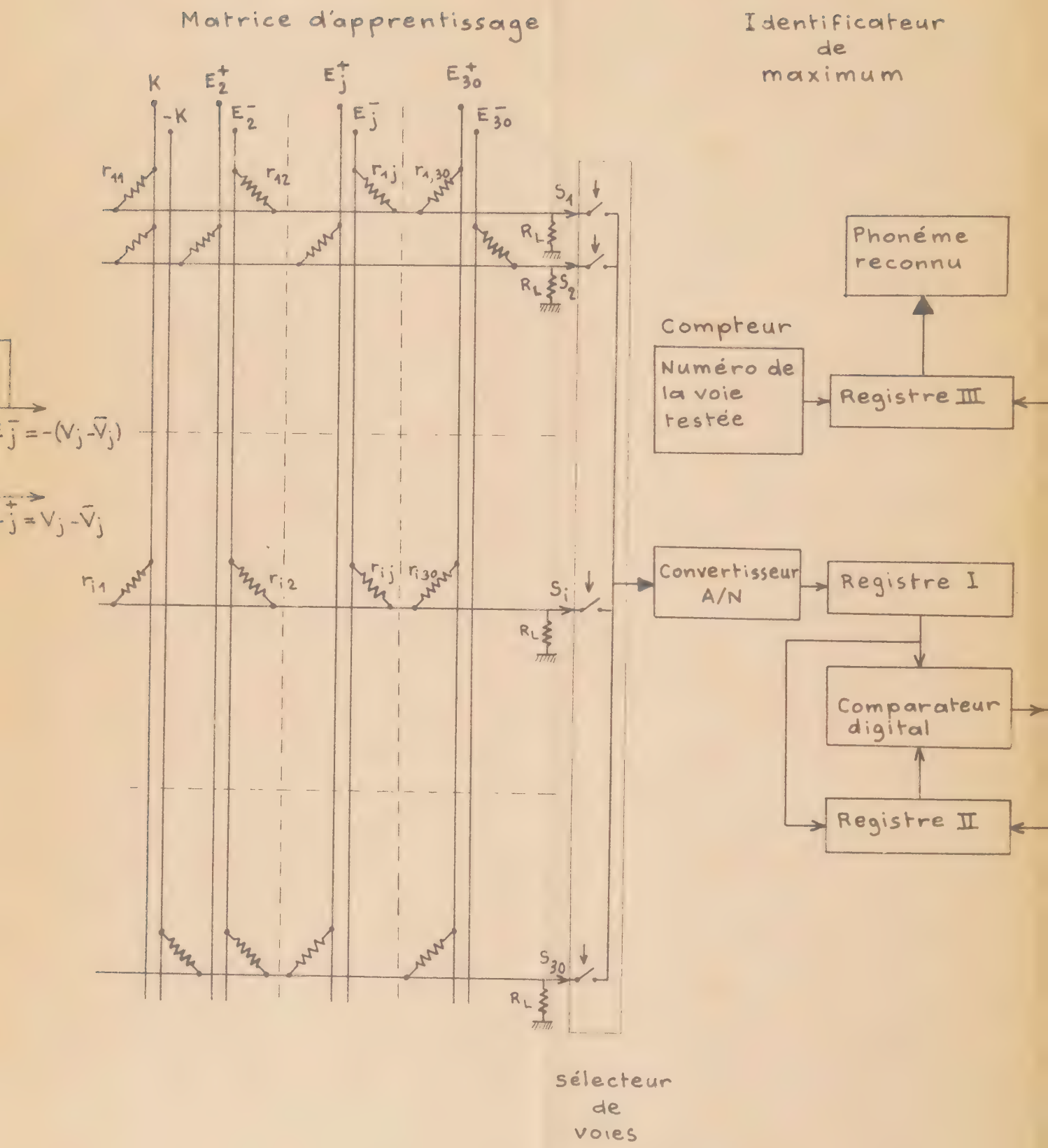


Schéma synoptique du systéme de reconnaissance réalisé

Planche 4

l'un essentiellement numérique, organise en séquence les opérations de calibrage des signaux d'entrée des trois groupes d'intégrateurs définis dans le fonctionnement général, l'autre, entièrement analogique, génère les tensions E_j^+ et E_j^-.

II.2.1. Calibrage des signaux d'entrée des intégrateurs.

Cet ensemble est schématisé planche 5 tandis que le diagramme des signaux générés dans l'ensemble lors d'un échantillonnage est représenté planche 6.

Les deux bases de temps H_{100} et H qui commandent l'évolution des opérations dans le système sont totalement indépendantes.

La première, de fréquence 100 Hz, détermine la cadence d'échantillonnage et, en particulier ici, ordonne la saisie des données issues du prétraitement, par le registre R.

La deuxième, de fréquence plus élevée, 6,4 kHz, commande les décalages dans le registre et les calibrages en temps des impulsions r_1, r_2, r_3 et S_f.

Ces quatre impulsions sont formées à l'aide de quatre monostables déclenchables, soit sur le flanc croissant (entrée (M)), soit sur le flanc décroissant (entrée (D)) d'un signal d'entrée. La largeur des impulsions de sortie d'un monostable qui donne également la sortie complémentée, correspond au temps de déséquilibre de l'oscillateur et dépend donc de la constante de temps de celui-ci.

Pour des raisons inhérentes au cablage, le mot binaire présent sur les sorties du prétraitement est en fait le complément logique du mot caractéristique de la forme vocale à l'instant d'échantillonnage considéré. Le transfert parallèle de ce mot dans le registre R est validé par les signaux complémentaires décalés C_R et P_R (voir planche 6) ; il se termine sur le flanc descendant de P_R.

Les opérations de calibrage commencent alors sur les entrées de chaque groupe d'intégrateurs.

II.2.1.1. Calibrage C 1 .

Cette opération consiste à élaborer le signal d'entrée des 24 premiers intégrateurs et débute par le déclenchement du monostable 1 , sur le flanc descendant de P_R .

L'impulsion $\overline{r_1}$, de largeur θ_1 , ainsi générée est appliquée sur une des entrées de 24 portes NOR , l'autre entrée de chaque porte étant une des sorties du registre R.

L'équation générale de la sortie des portes s'écrit par conséquent

$$s^*_j = \overline{\overline{s_j} + \overline{r_1}} = s_j \cdot r_1 \quad ;$$

elle traduit que le passage d'une impulsion calibrée r_1 vers l'entrée de l'intégrateur j est uniquement validé quand $s_j = 1$.

Durant cette opération, l'horloge H est inopérante sur l'entrée H_d de commande du décalage série dans le registre R (la porte I est alors verrouillée par $\overline{r_1}$) et de plus, les compteurs I et II sont remis à zéro.

II.2.1.2. Calibrage C 2.

Lorsque l'impulsion r_1 retombe au niveau 0 , la porte I se débloque et le signal d'horloge H , restitué sur l'entrée H_d , provoque le décalage (vers le bas sur le schéma) du mot binaire stocké dans le registre.

Les huit premières impulsions d'horloge ordonnent la sortie série du premier, du deuxième et du troisième octet de ce mot, respectivement sur les bits 9 , 17 et 25 du registre, le décalage d'un bit vers la position suivante s'effectuant sur le front de montée d'une impulsion.

Ce même train d'impulsions déclenche huit fois de suite à la descente le monostable 2 ; il est restitué sur la sortie de l'oscillateur sous la forme d'un train de huit impulsions calibrées $\overline{r_2}$ de largeur θ_2 (la porte II est durant cette phase déverrouillée). La largeur de $\overline{r_2}$ est compatible avec la fréquence d'horloge, c'est-à-dire que θ_2 est inférieure à θ',

intervalle de temps délimité par le flanc décroissant d'une impulsion d'horloge et le flanc montant de la suivante.

L'élaboration du signal d'entrée des intégrateurs 25 , 26 et 27 s'effectue dans trois portes NOR à deux entrées : la première entrée reçoit en série un des trois octets, la deuxième le train d'impulsions calibrées. Ainsi, dans chaque porte, la $k^{ième}$ impulsion $\bar{r}_2$ est validée ou non par le $k^{ième}$ bit de l'octet ; en d'autres termes, lors du décalage, il y a autant d'impulsions r_2 intégrées que de bits à 1 dans chaque octet.

Cette opération doit se terminer après le huitième décalage dans le registre ; aussi, le passage de la huitième impulsion d'horloge, par le basculement de 0 à 1 du digit de poids 3 du compteur I, déclenche à la montée le monostable 4, forçant la sortie de la porte II au niveau logique 0 durant le déséquilibre de celui-ci. Ce déséquilibre, de durée θ_4 , est suffisamment long pour inhiber le déclenchement du monostable 2 tant que se poursuivra le décalage dans le registre.

II.2.1.3. Calibrage C 3.

Cette opération s'applique aux intégrateurs 28 , 29 et 30, après le test du nombre N de bits à 1 du mot binaire stocké, en début d'échantillonnage, dans le registre R .

L'affectation des entrées s^*_{28} , s^*_{29} et s^*_{30} de ces intégrateurs aux différents tests est la suivante :

s^*_{28} est affectée à $N > 7$

s^*_{29} est affectée à $N = 0$

s^*_{30} est affectée à $0 < N \leq 7$

Le nombre N est élaboré dans le compteur II lors du décalage. Le signal d'incrémentation H2 de ce compteur correspond au mot binaire généré en série sur la 25ème sortie du registre et que l'on a précédemment mis en forme dans la porte III, par les impulsions d'horloge H synchronisant le décalage ($H2 = H \cdot s_{25}$).

La nature des tests effectués sur N suscite deux remarques :

- si lors du glissement du mot dans le registre, N devient supérieur à 7 , il est inutile de poursuivre le décalage ;

- si, par contre, N reste inférieur à 7, le décalage doit s'opérer jusqu'à la sortie complète du mot.

Le premier cas s'illustrepar le basculement de 0 à 1 du digit 2^3 du compteur II , le deuxième, par un basculement identique de la sortie du décodage de la position 25 du compteur I.

Dans les deux cas, le résultat obtenu est le verrouillage de la porte IV et consécutivement celui de l'horloge H dans la porte I . Le décalage est alors arrêté dans le registre, puisque l'entrée H_d est maintenue au niveau logique 0.

On peut conclure de ce qui précède, que le nombre d'impulsions d'horloge appliquées en H_d est compris entre 8 et 25.

Les trois tests du nombre N sont ensuite réalisés en parallèle. La réponse à chacun de ces tests est composée respectivement dans trois portes avec l'impulsion calibrée(r_3 ou $\overline{r}_3$) de largeur θ_3 que l'on a générée en déclenchant le monostable 3 par le front de descente résultant du verrouillage de la porte IV ; ce front déclenche également le monostable 5. Une seule réponse aux tests étant évidemment positive, le passage de l'impulsion r_3 est validé vers l'entrée d'un seul des trois intégrateurs.

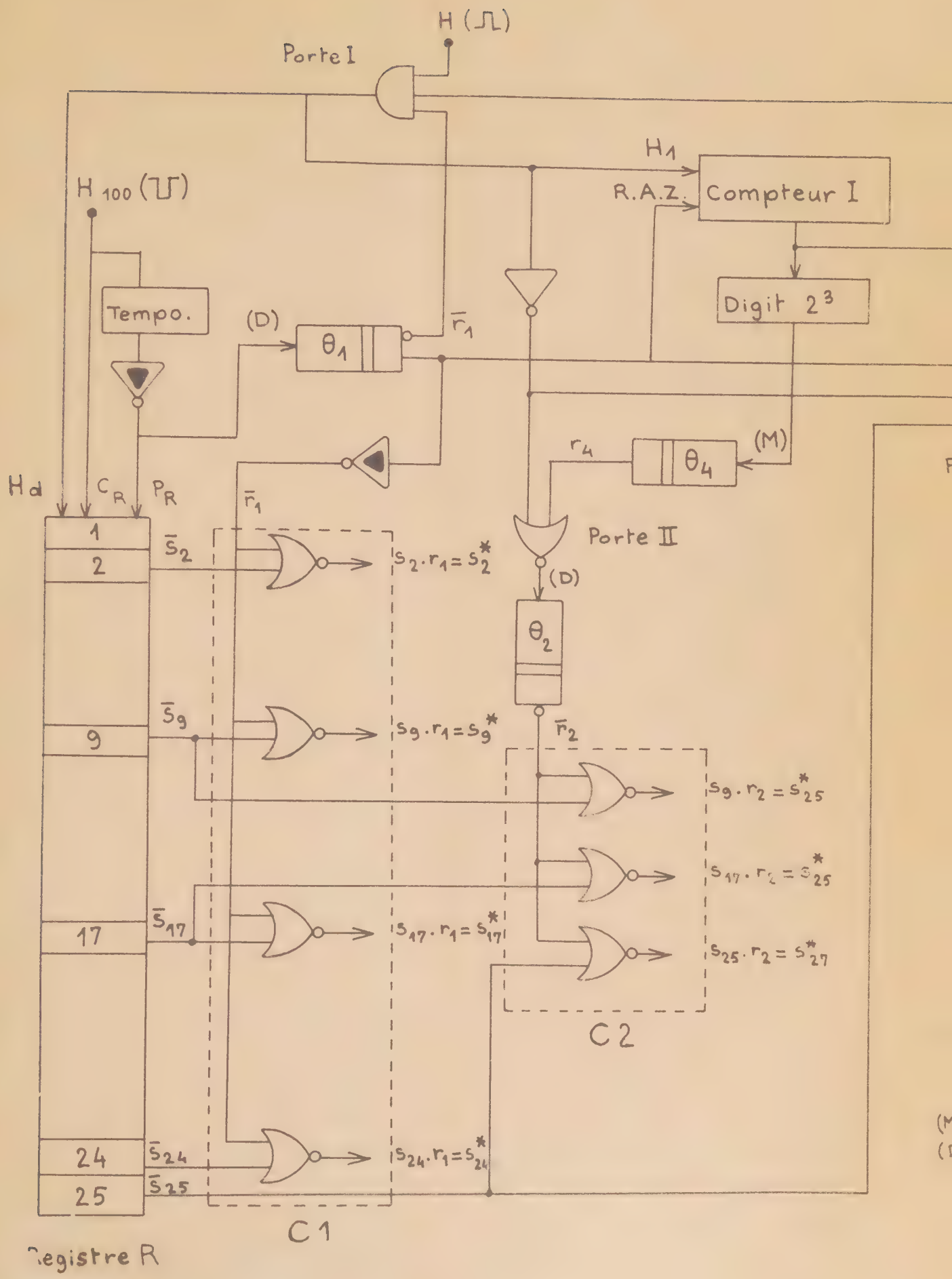
Les équations logiques de la sortie des portes s'écrivent :

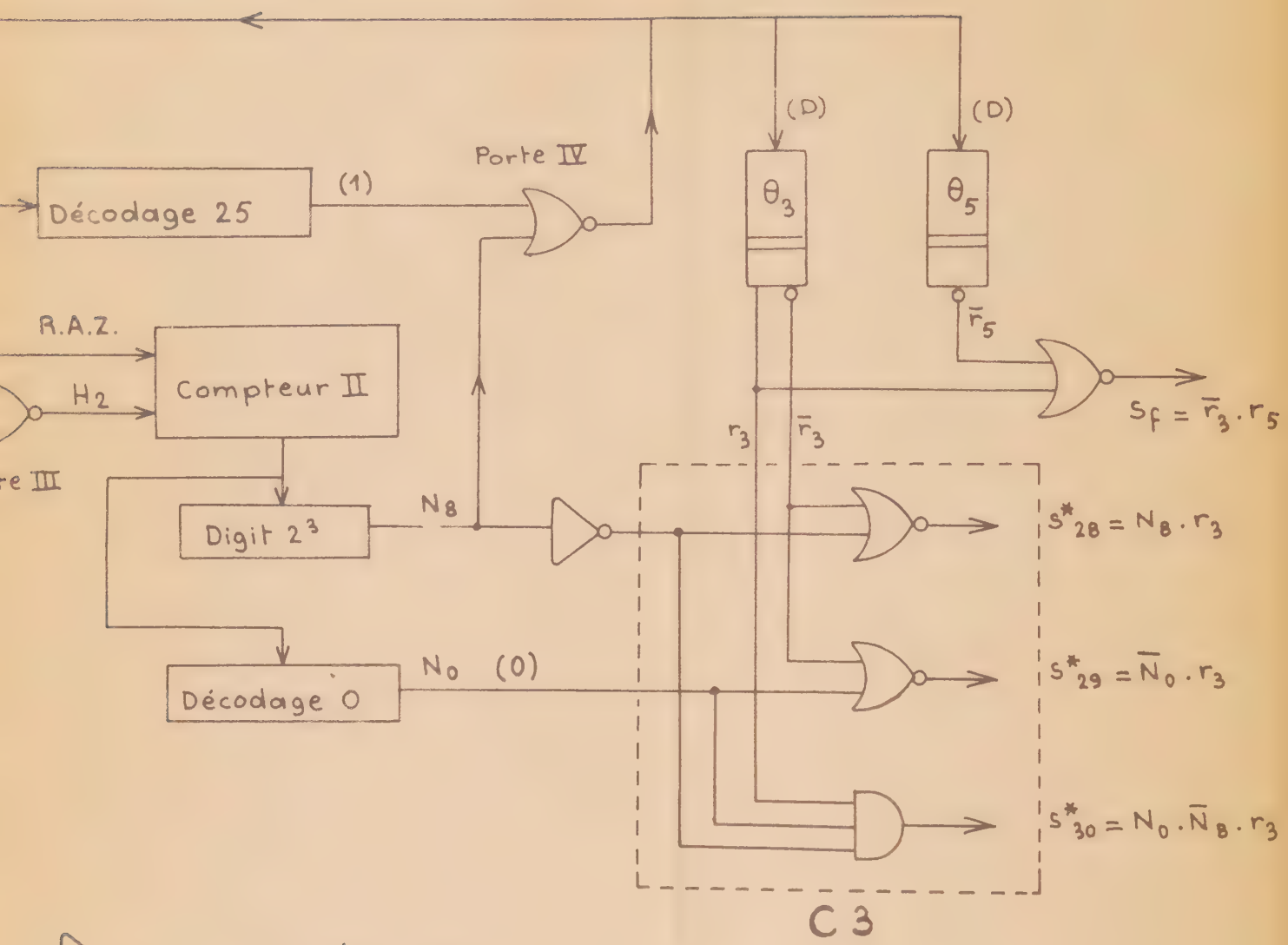
$$s^*_{28} = N_8 . r_3 \quad \text{avec} \quad \begin{cases} N_8 = 1 & \text{si } N > 7 \\ N_8 = 0 & \text{si } N \leqslant 7 \end{cases}$$

$$s^*_{29} = \overline{N}_o . r_3 \quad \text{avec} \quad \begin{cases} N_o = 0 & \text{si } N = 0 \\ N_o = 1 & \text{si } N \neq 0 \end{cases}$$

$$s^*_{30} = N_o . \overline{N}_8 . r_3 \text{ avec , d'après ci-dessus}$$

$$\begin{cases} N_o . \overline{N}_8 = 1 & \text{si } 0 < N \leqslant 7 \\ N_o . \overline{N}_8 = 0 & \text{si } N = 0 \text{ ou } N > 7 \end{cases}$$


H (⎍)
Porte I
H1
R.A.Z.
Compteur I
H 100 (⎍)
Digit 2³
Tempo.
(D)
θ1
r̄1
r4
θ4
(M)
Hd
CR
PR
Porte II
(D)
θ2
r̄2
1
2
9
17
24
25
s̄2
s̄9
s̄17
s̄24
s̄25
s2.r1 = s*2
s9.r1 = s*9
s17.r1 = s*17
s24.r1 = s*24
s9.r2 = s*25
s17.r2 = s*25
s25.r2 = s*27
C1
C2
Registre R

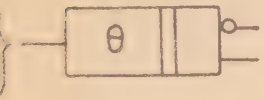


Complémenteur

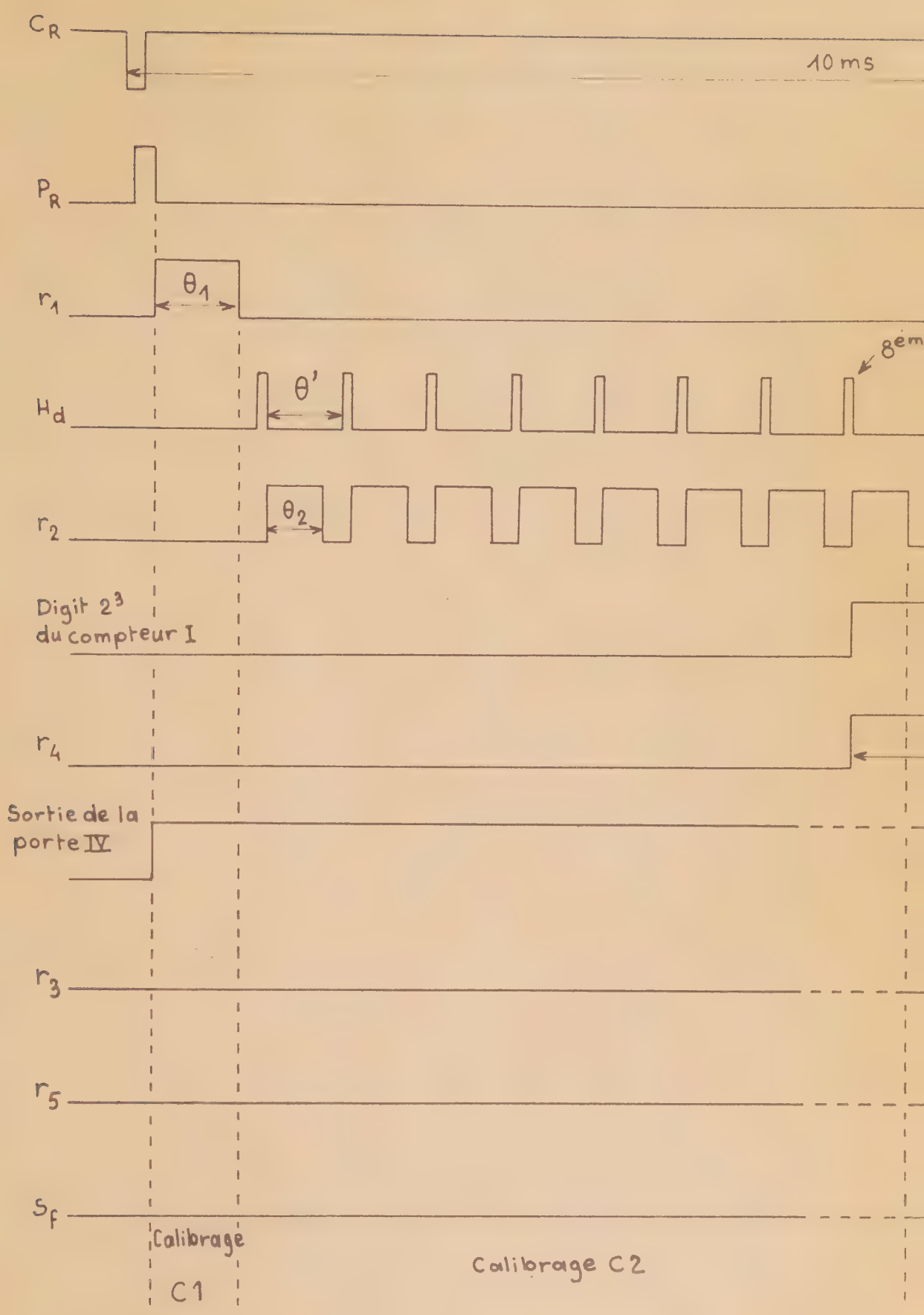
Buffer complémenteur

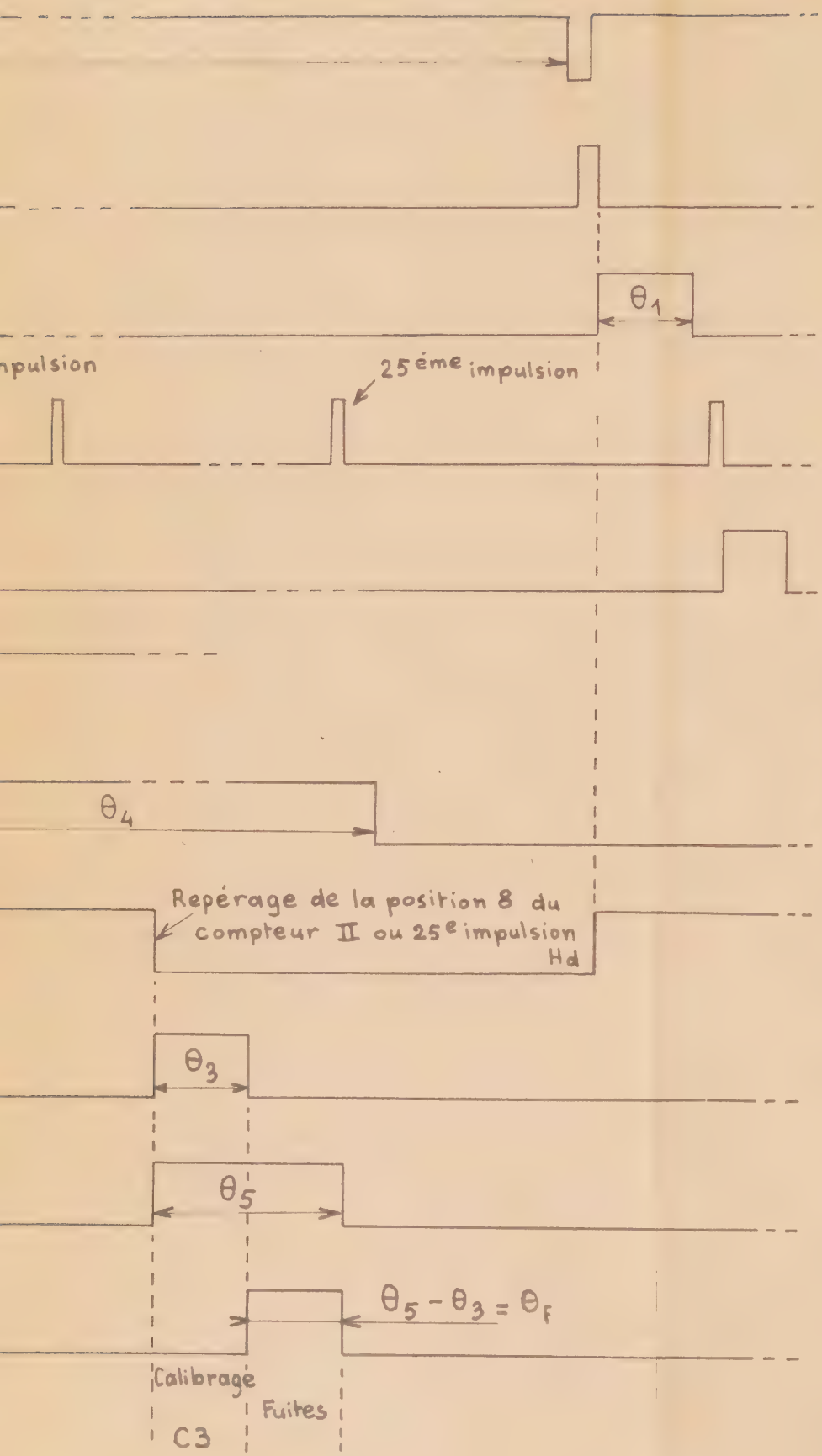
Circuit NOR (OU-NON)

Circuit ET 3 entrées



$\bar{Q}$ (⊔) Monostable de retard θ déclenchable
Q (⊓) à la montée (M) ou à la descente (D)


C_R
10 ms
P_R
r_1
θ_1
H_d
θ'
8ème
r_2
θ_2
Digit 2^3
du compteur I
r_4
Sortie de la
porte IV
r_3
r_5
S_f
Calibrage
C1
Calibrage C2



Fréquence de Hd = 6,4 kHz
$\theta_1 = 625\,\mu s$
$\theta_2 = 125\,\mu s$
$\theta_3 = 625\,\mu s$
$\theta_4 = 2{,}8\ ms$
$\theta_5 = 666{,}7\,\mu s$
$\theta_f = \theta_5 - \theta_3 = 51{,}7\,\mu s$

Planche 6

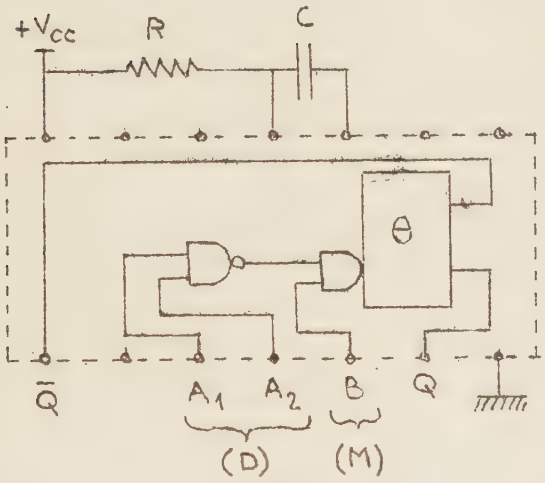
La dernière opération de calibrage est l'élaboration du signal S_f de commande des fuites. Ce signal résulte de la composition dans la porte V des impulsions r_3 et $\overline{r}_5$ de largeur respective θ_3 et $\theta_5 > \theta_3$; il a pour équation $S_f = \overline{r}_3 . r_5$ et pour largeur $\theta_f = \theta_5 - \theta_3$.

II.2.1.4. Réalisation à l'aide de circuits intégrés.

La réalisation de l'ensemble logique que l'on vient de décrire ainsi que tous ceux qui le seront par la suite, est faite à l'aide de circuits intégrés logiques TTL de la série SN 74 Texas Instruments utilisant l'intégration à faible échelle (SSI) pour les fonctions logiques simples (opérateurs, bascules, monostables) et à moyenne échelle (MSI) pour les fonctions plus complexes (registres, compteurs,...).

Monostable : (SN 74121 N)

Le monostable est intégré dans un boîtier comme l'indique la figure 1, le réseau RC de temporisation étant extérieur.



Représentation simplifiée.

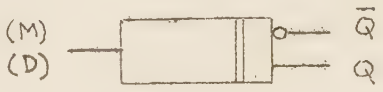


Figure 1.

A_1 et A_2 sont les entrées (D) de déclenchement au front de descente (B doit alors être au niveau 1) et B, l'entrée (M) de déclenchement à la montée (avec A_1 ou $A_2 = 0$).

La largeur de l'impulsion générée sur la sortie Q est

$$\theta = RC \ \text{Log}\ 2$$

On fera ultérieurement le calcul de la largeur des diverses impulsions calibrées définies dans l'ensemble.

Un dispositif de compensation interne au boîtier rend la largeur de l'impulsion indépendante de V_{cc} et de la température ; la stabilité de l'impulsion est seulement limitée par la précision des composants extérieurs qui déterminent la constante de temps.

Horloges H_{100} et H .
. . . . . .

Chaque horloge est un générateur d'impulsions rectangulaires, réalisé en connectant en série deux monostables, dont la sortie Q' du second est réinjectée sur l'entrée B du premier (figure 2).

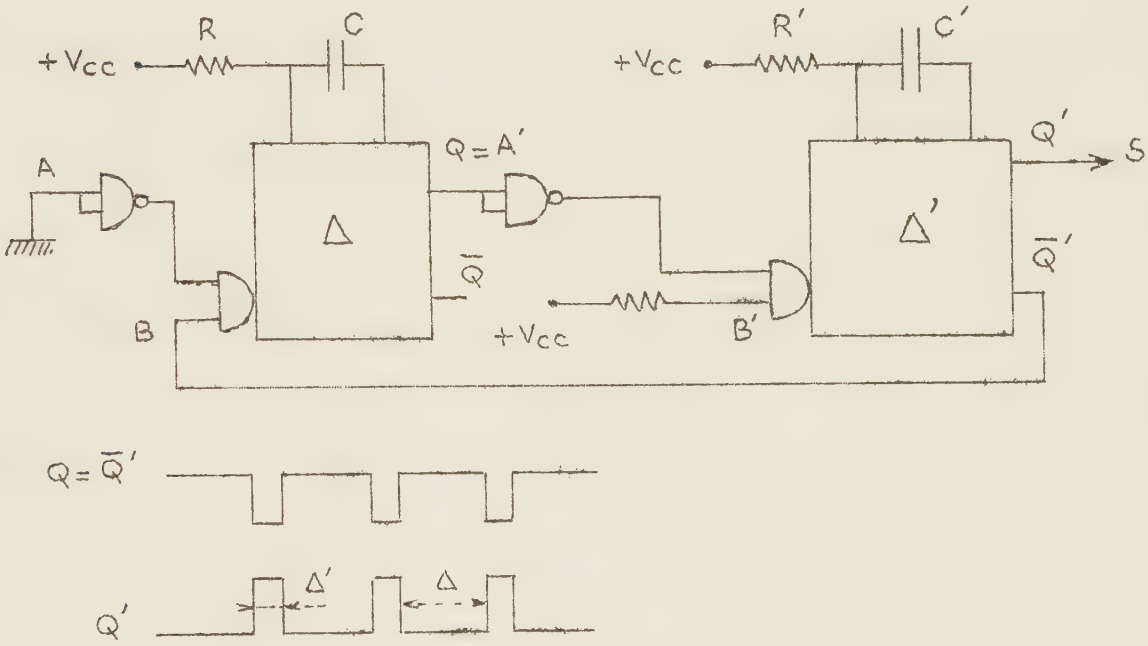


Figure 2.

On obtient sur la sortie Q' des créneaux ajustables en durée (Δ') et en espacement (Δ).

L'horloge H_{100} est, plus exactement, formée de ressauts (ce sont des créneaux complémentés), pris sur la sortie $\overline{Q}'_1$ d'un premier montage et apparaissant toutes les 10 ms. Si par conséquent on se fixe $\Delta'_1 = 10\,\mu s$:

* $\Delta'_1 = R'_1\,C'_1 \,\mathrm{Log}\,2 = 10\,\mu s$ entraîne $R'_1\,C'_1 = 14{,}3\,\mu s$,

constante de temps affichée à l'aide d'une résistance R'_1 ajustable de 2,2 kΩ et d'une capacité C'_1 de 10 nF)

$$* \quad \Delta_1 = 10000 - \Delta'_1 = R_1 C_1 \text{ Log } 2 \quad \text{donne } R_1 C_1 = 14{,}27 \text{ ms}$$

(R_1 résistance ajustable de 220 kΩ et C_1 de 100 nF)

L'horloge H de fréquence 6,4 kHz est obtenue en divisant par deux le nombre d'impulsions générées par une base de temps H' de fréquence 12,8 kHz (l'utilisation de cette base de temps H' sera spécifiée dans la dernière partie de ce travail). Le montage réalisé est celui représenté sur la figure 3.

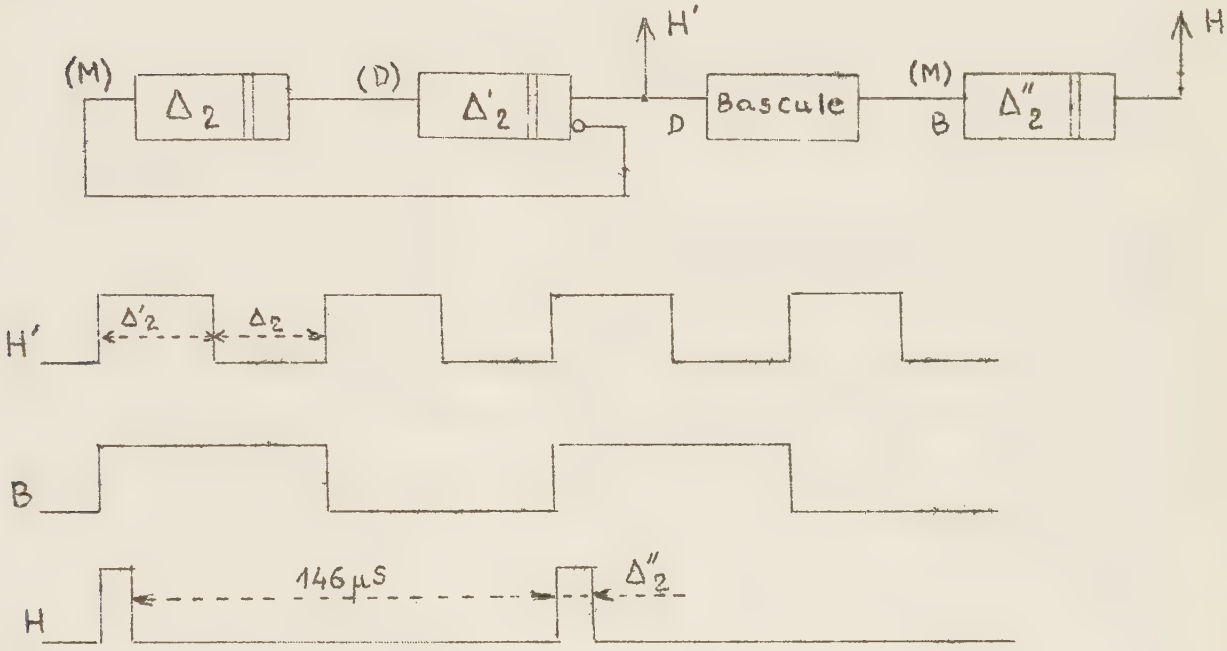


Figure 3.

On a

$$\Delta_2 + \Delta'_2 = \frac{1}{12{,}8.10^3} = 78 \mu s$$

où on impose

$$\Delta_2 = \Delta'_2 = R_2 C_2 \text{ Log } 2 = R'_2 C'_2 \text{ Log } 2 = 39 \mu s$$

ce qui fixe

$$R_2 C_2 = R'_2 C'_2 = 55{,}7 \mu s$$

($R_2 = R'_2$ résistances ajustables de 4,7 kΩ et $C_2 = C'_2 = 22$ nF).

La configuration du signal H est motivée par les opérations de calibrage C2 (cf. paragraphe II.2.1.2.), nécessitant l'intégration d'une impulsion r_2 entre deux décalages dans le registre R ; sa période est égale à $4\Delta_2$ (156 μs), ce qui donne pour le dernier monostable, Δ''_2 étant fixée à 10 μs :

$$\Delta''_2 = R''_2 C''_2 \text{ Log } 2 = 10 \mu s \qquad \text{donc} \qquad R''_2 C''_2 = 14,3 \mu s$$

(R''_2 résistance ajustable de 2,2 kΩ et C''_2 = 10 nF).

Registre R.

Ce registre, de dimension 25 bits, est composé d'une cascade de 5 registres à 5 bits SN 7496 N fonctionnant en entrée série et parallèle, sortie série et parallèle.

La prédétermination du registre s'opère lorsque les entrées de forçage C_R et P_R sont respectivement aux niveaux 0 et 1 ; le décalage y est validé lorsque $C_R = 1$ et $P_R = 0$. En fait, on applique en C_R et P_R deux impulsions complémentaires légèrement décalées qui permettent de remettre le registre à zéro avant sa prédétermination, le décalage des deux impulsions étant formé dans un réseau RC de temporisation.

Compteurs I et II.

Ils sont constitués par des compteurs binaires SN 7493 N et ont respectivement pour dimension 5 et 4 bits.

Les décodages qui leur sont associés sont composés dans des portes ET. Ainsi, si l'on définit par E, D, C, B et A les bits des compteurs dans l'ordre décroissant des poids, on obtient comme équations de décodage :

- de la position 25 sur le compteur I :

$$\text{décodage } 25 = E.D.\overline{C}.\overline{B}.A$$

La sortie du décodeur est à l'état binaire 1 pour un décodage vrai.

- de la position 0 sur le compteur II :

$$\text{décodage } 0 = N_o = \overline{\overline{D}.\overline{C}.\overline{B}.\overline{A}}$$

l'état binaire 0 étant affecté au décodage vrai.

II.2.2. Elaboration des entrées E_j^+ et E_j^- de la matrice d'apprentissage. [9 , 10]

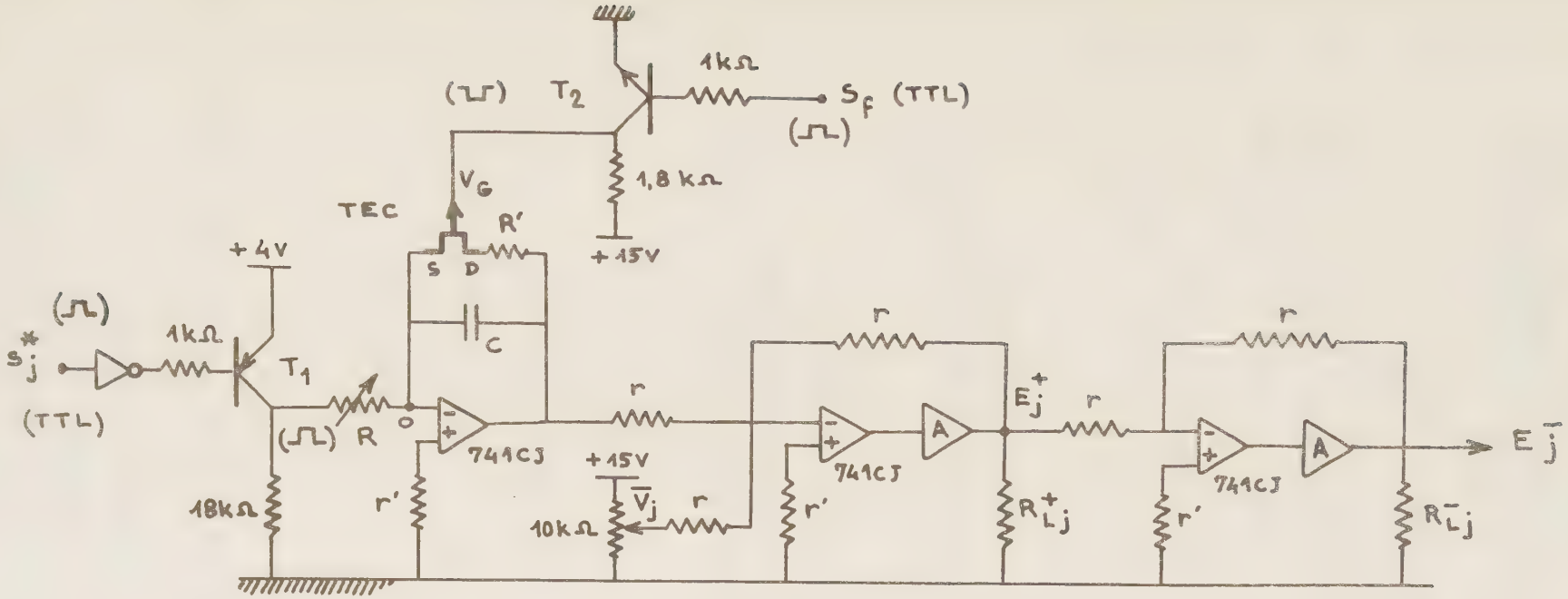
L'ensemble analogique qui génère les tensions E_j^+ et E_j^- appliquées sur les entrées de la matrice d'apprentissage est constitué de 30 chaînes identiques dont les entrées sont les signaux s_j^* élaborés et synchronisés par l'ensemble numérique décrit précédemment. Le schéma d'une chaîne analogique est donné planche 7.

Les opérations effectuées sont : le calibrage des impulsions r du signal s_j^* en amplitude, l'intégration de ces dernières, l'application des fuites sur la capacité de bouclage de l'intégrateur, l'élaboration de l'écart $E_j^+ = V_j - \overline{V}_j$ entre la tension de sortie de l'intégrateur et la valeur moyenne $\overline{V}_j$ de cette tension, enfin, l'élaboration de E_j^- opposée de E_j^+.

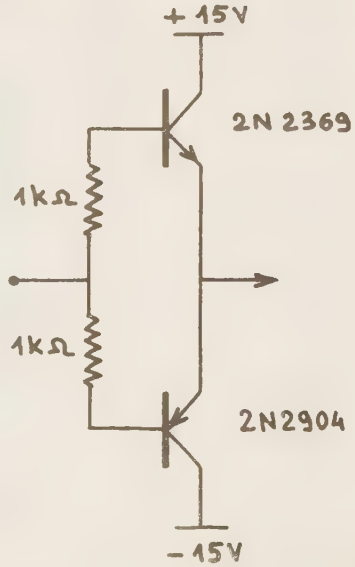
Les impulsions r , calibrées en amplitude (niveau bas : 0 V, niveau haut : 4 V), sont formées sur le collecteur du transistor T_1 , monté en émetteur commun et fonctionnant en commutation.

La capacité C de contre-réaction de l'amplificateur opérationnel qui fait suite, est pontée par un transistor à effet de champ (T. E. C. canal P) associé en série à une résistance R'.

Le T. E. C. joue ici le rôle d'un interrupteur analogique ; en position fermée, si la grille du transistor est portée à un potentiel $V_G = 0V$, en position ouverte si $V_G = 15$ V. La tension de grille V_G est élaborée sur le collecteur du transistor de commutation T_2 dont la base est commandée par le signal TTL S_f de validation des fuites.



Amplificateur de puissance



Composants

TEC	P 1086
T1	2N 2904
T_2	2N 2369
R	10 kΩ ajustable
R' #	1750 Ω
r	100 kΩ
r'	50 kΩ
C	100 nF

Elaboration des entrées E_j^+ et E_j^- de la matrice d'apprentissage

Planche 7

Ce montage fonctionne de la manière suivante :

- lorsqu'une impulsion r se présente sur l'entrée, le transistor à effet de champ est bloqué ($S_f = 0$), l'impédance connectée en parallèle sur C infinie, et le montage se comporte comme un intégrateur pur.

- lorsque l'impulsion disparaît, l'entrée est à un potentiel nul ; la tension de sortie de l'intégrateur reste constante jusqu'à ce que S_f passe au niveau logique 1 et provoque le débloquage du T. E. C. ; la capacité C se décharge alors dans R' tant que S_f reste au niveau haut.

Etudions le comportement de la tension de sortie $-V_j$ de l'amplificateur au cours de ces deux phases, en supposant :

- qu'au temps $t = 0$, la capacité est totalement déchargée ($-V_j=0$)

- que l'impulsion r , de largeur θ est appliquée sur l'entrée entre les instants $t = 0$ et $t = t_1$ et l'impulsion S_f , de largeur θ_f , entre t_2 et t_3.

Soient RC la constante de temps de l'intégrateur et a le potentiel correspondant au niveau haut d'une impulsion r . La tension de sortie $-V_j$ durant l'intégration a pour expression :

$$-V_j = -\frac{1}{RC}\int a\,dt = -\frac{at}{RC} + K ,$$

(avec $K = 0$ puisque $V_j^o = 0$)

rampe de pente $-\frac{a}{RC}$ (voir figure 4).

En particulier, à la fin de l'intégration, ce potentiel vaudra

$$-V_j^1 = -\frac{a\,t_1}{RC} = -\frac{a\theta}{RC}$$

puis restera constant jusqu'à l'instant $t = t_2$.

Lorsque le T. E. C. est saturé, sa résistance drain (D) - source (S) se limite à la résistance R_{0N} (de l'ordre de 50 Ω pour le T. E. C. utilisé) et la capacité C se décharge dans les résistances R' et R_{0N} en série, cette dernière étant mise à la masse virtuelle au point O.

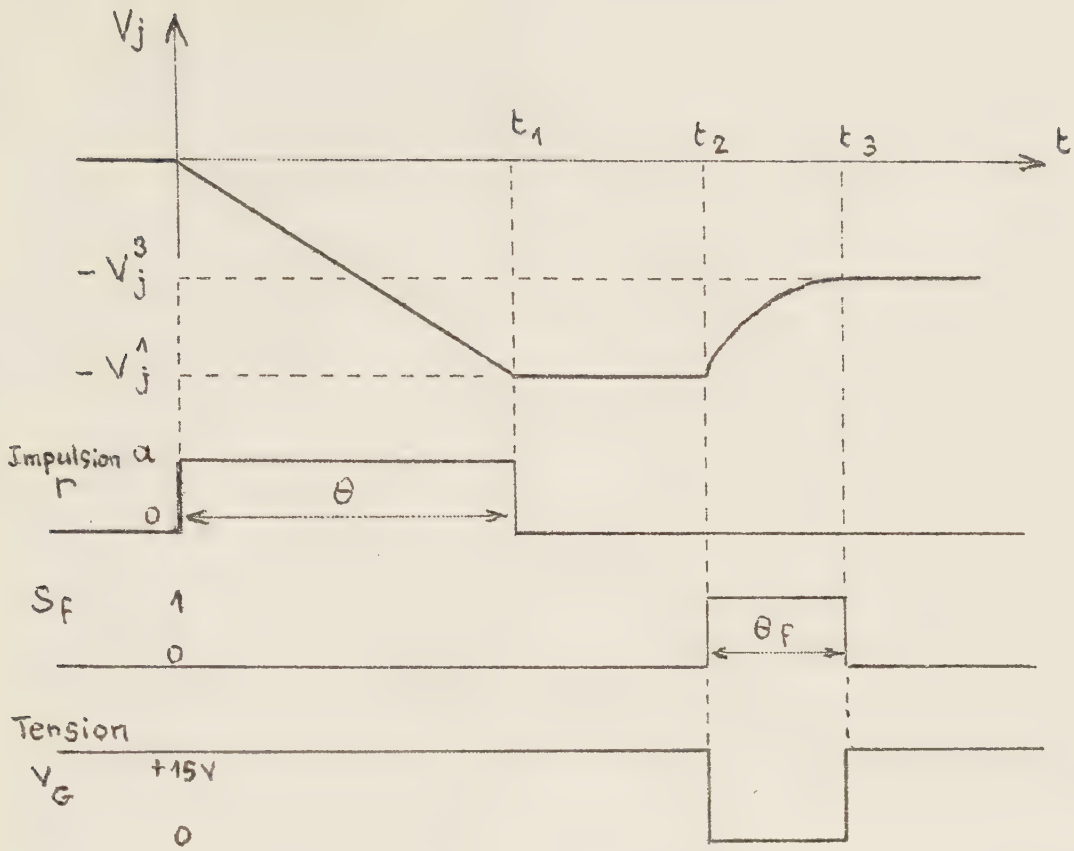


Figure 4.

Soit i le courant circulant dans le circuit de contre-réaction durant la décharge ; on peut écrire :

$$(R' + R_{0N})\, i + \int \frac{i\, d\, t}{C} = 0 \qquad (1)$$

Sachant que la charge q de la capacité est liée à i par la relation :

$$- q = \int i\, d\, t = - C \,.\, V_j$$

on obtient

$$i = - \frac{dq}{dt} = - C \frac{d\, V_j}{dt}$$

et l'équation (1) devient, si l'on pose $R' + R_{0N} = R_f$

$$R_f\, C \frac{d\, V_j}{dt} + V_j = 0 \qquad (2)$$

La solution de l'équation (2) est de la forme

$$- V_j = K' . e^{-t/R_f C}$$

avec comme condition initiale $-V_j = -\frac{a\theta}{RC} = -V_j^1$ quand $t = t_2$.

L'expression finale de la tension de sortie lors de la décharge est donc

$$-V_j = -V_j^1 \, e^{-\frac{(t-t_2)}{R_f \cdot C}}$$

et en particulier à la fin de celle-ci ($t = t_3$) :

$$-V_j = -V_j^1 \, e^{-\frac{(t_3-t_2)}{R_f \cdot C}} = -V_j^1 \, e^{-\frac{\theta_f}{R_f \cdot C}} = -V_j^3$$

Puis la tension reste constante jusqu'à l'échantillonnage suivant.

Dans le cas envisagé ci-dessus, on a étudié les variations du potentiel de sortie d'un intégrateur, lorsqu'une seule impulsion r est intégrée durant le pas d'échantillonnage, ce qui est le cas des chaînes indicées de 1 à 24 et de 28 à 30.

Pour les chaînes 25, 26 ou 27, il y aura autant d'intégrations que de bits à 1 dans l'octet.

On donne sur la figure 5 l'exemple du mode d'intégration de l'octet 10111011 présenté en série sur l'entrée d'une de ces chaînes.

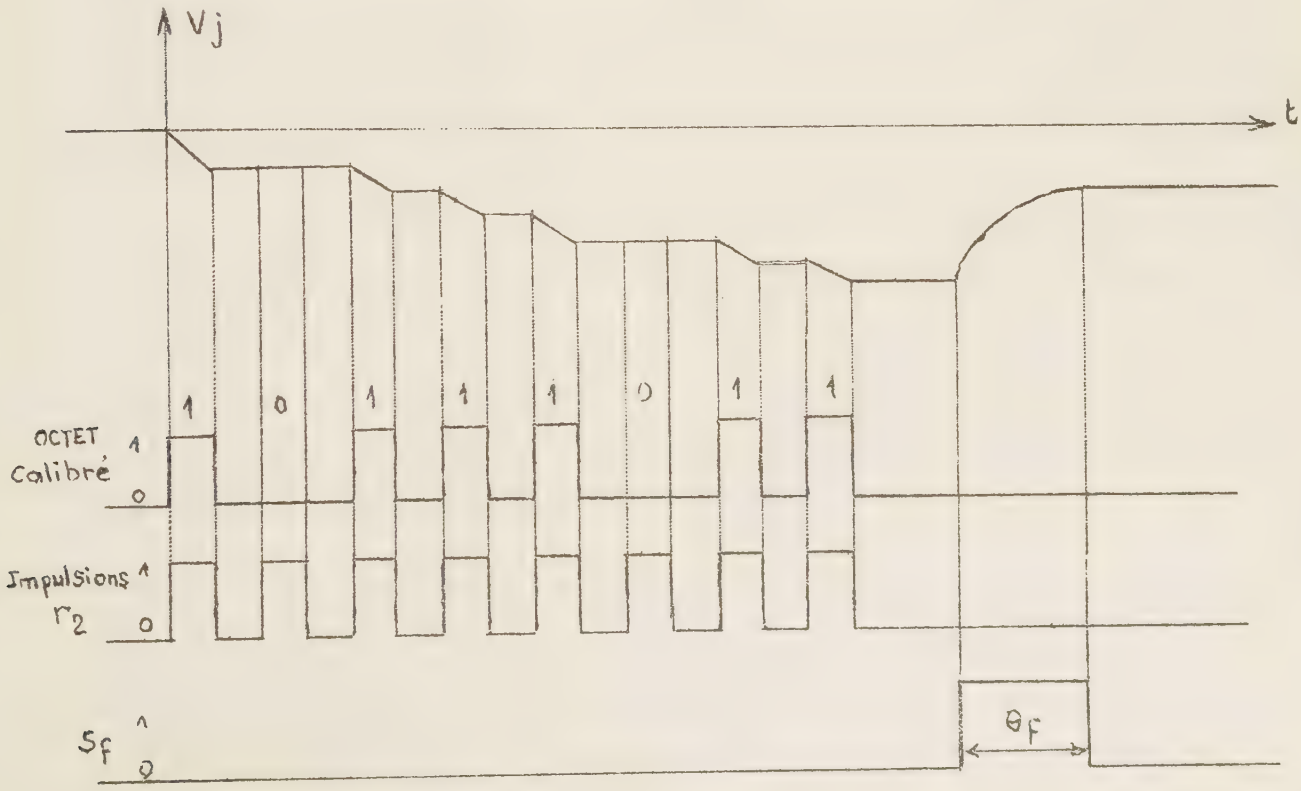


Figure 5.

La tension $-V_j$ est ensuite sommée à la tension moyenne $\bar{V}_j$ dans un sommateur - inverseur de gain unité pour former l'écart $E_j^+ = V_j - \bar{V}_j$.

La valeur $\bar{V}_j$ est affichée à l'aide d'un potentiomètre, l'entrée s_j^* étant court-circuitée à la masse et la capacité C totalement déchargée. On règle alors le potentiomètre jusqu'à ce qu'apparaisse en E_j^+ le potentiel $-\bar{V}_j$.

La grandeur E_j^- opposée de E_j^+, est formée dans un inverseur simple de gain unité ; on a $E_j^- = -E_j^+ = -(V_j - \bar{V}_j)$.

Lorsque les tensions E_j^+ et E_j^- sont appliquées sur les deux entrées de la $j^{ième}$ colonne de la matrice d'apprentissage, les sorties des sommateurs sont chargées par les impédances d'entrée de la matrice (respectivement R_{Lj}^+ et R_{Lj}^-) et les amplificateurs doivent être capables de débiter des courants suffisants (environ 10 mA).

Comme ce n'est pas le cas pour les amplificateurs utilisés, on a placé dans la boucle directe des sommateurs, un amplificateur de puissance A, qui est un push-pull série à deux transistors complémentaires et émetteurs couplés (voir planche 7). Cet étage de puissance s'assimile, pour le transistor passant, à un collecteur commun dont la résistance d'émetteur correspond à la résistance émetteur-collecteur en bloqué du deuxième transistor ; son impédance de sortie est, par conséquent, très faible (de l'ordre de 40 Ω).

On peut en conclure que l'impédance de sortie d'un sommateur est encore beaucoup plus faible que celle du push-pull, puisque notablement diminuée par la contre-réaction de l'amplificateur. Cette remarque s'avère importante à noter pour le calcul des résistances de la matrice, car les deux montages sommateur - inverseur se comportent alors comme deux générateurs de tension pratiquement parfaits.

On terminera la description d'une chaîne analogique en soulignant que l'entrée non inverseur des amplificateurs opérationnels employés, est fixée à la masse à travers une résistance r', afin de compenser le déséquilibre entre les deux entrées ; ce déséquilibre est minimal pour :

$$r' = \frac{r.r}{r+r} = \frac{r}{2}$$

II.2.3. Calcul des paramètres intervenant dans l'élaboration des entrées de la matrice d'apprentissage.

II.2.3.1. Position du problème.

Nous avons décrit la configuration de l'ensemble hybride qui élabore les entrées de la matrice d'apprentissage, sans nous préoccuper lors des explications de la valeur des constantes introduites; il reste maintenant à les calculer, en utilisant comme support les résultats obtenus par simulation sur ordinateur.

Il s'agit tout d'abord de préciser comment sont effectuées les opérations numériques sur une voie durant la reconnaissance simulée, la séquence de ces opérations étant schématisée planche 8 :

- l'intégration se définit par la relation

$$I_t = I'_{t-1} + k \qquad \text{avec} \qquad k = 20 \tag{3}$$

où k peut s'identifier à une impulsion calibrée r et I'_{t-1} à l'état de sortie de l'intégrateur, à la fin des fuites appliquées lors de l'échantillonnage $t-1$;

- les fuites sont simulées en retranchant à I_t une valeur qui lui est proportionnelle

$$I'_t = I_t - \frac{I_t}{4} = \frac{3}{4} I_t \tag{4}$$

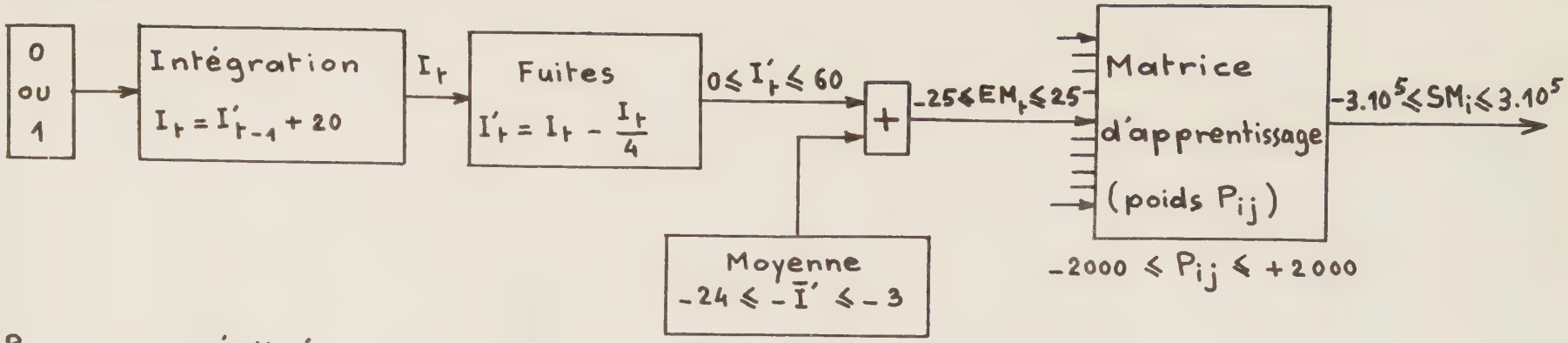
(le coefficient de proportionnalité a été choisi de façon à maximiser le taux de reconnaissance).

- et l'entrée EM_t de la matrice correspond à l'écart entre I'_t et sa valeur moyenne $\bar{I}'$:

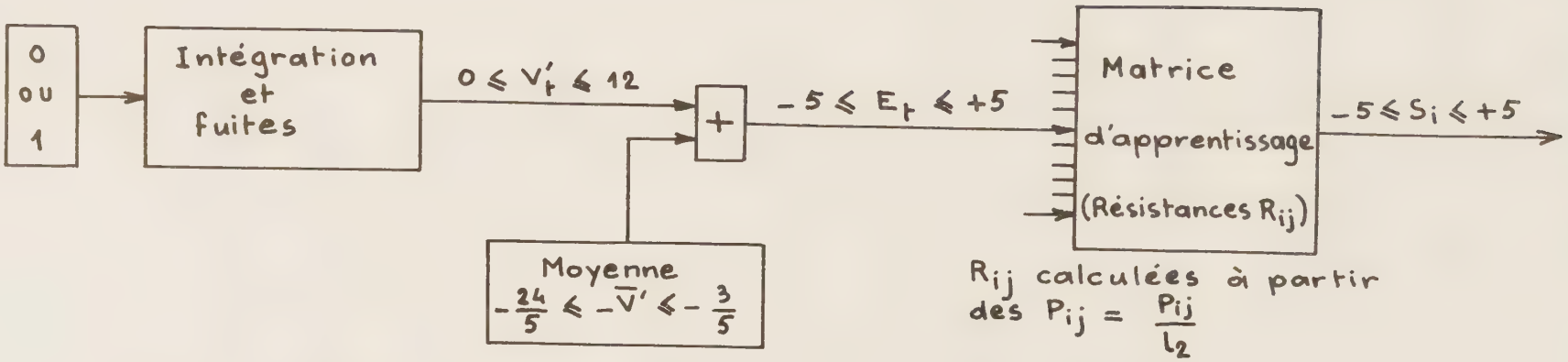
$$EM_t = I'_t - \bar{I}'$$

L'étude par simulation permet en particulier de chiffrer les valeurs extrêmes que peuvent prendre I'_t, EM_t et $\bar{I}'$; l'extremum le plus grand en module est par conséquent associé au potentiel le plus élevé au point correspondant dans la chaîne analogique et, compte tenu

Processus simulé :



Processus réalisé : les grandeurs ci-dessous sont exprimées en volts



$$l_1 = \frac{I'_t \max}{V'_t \max} = 5 \quad ; \quad l = \frac{SM_i \max}{S_i \max} = 6.10^4 \quad ; \quad l_2 = \frac{l}{l_1} = 12000$$

Correspondance entre les grandeurs simulées et les grandeurs réelles et notations utilisées

des contraintes technologiques, le rapport de ces deux valeurs détermine le coefficient de transposition liant une grandeur simulée à une grandeur réelle.

On convient ainsi d'associer à la valeur 60 de I'_t, valeur simulée maximale, le potentiel $V'_t = 12$ V sur la sortie de l'intégrateur avec fuites. Le coefficient de transposition vaut donc $\ell_1 = 5$ et avec $|EM_t| \leqslant 25$, les tensions d'entrée E_t de la matrice d'apprentissage varient entre $\pm$ 5V.

II.2.3.2. Calcul de la constante de temps RC d'un intégrateur.

Il est à remarquer que la valeur $I'_t = 60$ correspond à la limite vers laquelle tend I'_t quand une suite infinie de 1 se présente sur l'entrée de l'intégrateur. En effet, dans cette hypothèse, l'équation de définition de I'_t, lors du $t^{\text{ième}}$ échantillonnage, s'écrit, d'après (3) et (4) :

$$I'_t = \frac{3}{4}(I'_{t-1} + k) \qquad \text{avec} \qquad k = 20$$

En supposant qu'à l'instant initial l'intégrateur est remis à zéro ($I'_o = 0$), on obtient :

$$I'_1 = \frac{3}{4}k$$

$$I'_2 = \frac{3}{4}(I'_1 + k) = \frac{3}{4}k + (\frac{3}{4})^2 k$$

$$I'_3 = \frac{3}{4}(I'_2 + k) = \frac{3}{4}k + (\frac{3}{4})^2 k + (\frac{3}{4})^3 k$$

$$\vdots$$

$$I'_t = \frac{3}{4}(I'_{t-1} + k) = \frac{3}{4}k + (\frac{3}{4})^2 k + \ldots + (\frac{3}{4})^{t-1} k$$

On définit une série géométrique dont le $t^{\text{ième}}$ terme peut encore s'écrire :

$$I'_t = \frac{3}{4}\, k \,.\, \frac{(\frac{3}{4})^t - 1}{(\frac{3}{4}) - 1}$$

il a pour limite quand $t \to \infty$:

$$\lim_{t \to \infty} I'_t = 3\,k = 60$$

Ce résultat est important car il ne faut pas oublier que k correspond analogiquement, sur la sortie de l'intégrateur, au saut de tension provoqué par l'intégration d'une impulsion r de largeur θ (cf. paragraphe II.2.2.). Ainsi, compte tenu du coefficient de transposition ℓ_1, on a

$$\frac{a\theta}{RC} = \frac{k}{\ell_1} = 4 \text{ V}$$

où a est l'amplitude de l'impulsion calibrée et RC la constante de temps de l'intégrateur.

Avec a = 4 V, on obtient

$$RC = \theta$$

Pour calculer RC, il suffit donc de se fixer la largeur θ de l'impulsion intégrée ; mais il faut que ce choix soit compatible avec la période d'échantillonnage du signal vocal (10 ms).

Rappelons que l'élaboration des entrées de la matrice d'apprentissage sous-entend, dans le cas le plus défavorable :

- l'intégration d'une impulsion r_1 de largeur θ_1,
- 24 décalages dans le registre R à la fréquence de l'horloge avec 156 μs pour période d'horloge,
- l'intégration d'une impulsion r_3 de largeur θ_3,
- l'application de l'impulsion S_f de largeur θ_f

et que, durant le même pas d'échantillonnage, s'effectue l'identification du maximum.

On peut par exemple s'accorder 5 ms pour générer les impulsions r_1 et r_3 et pour décaler le contenu du registre ; ainsi, sachant que la durée θ n'est autre que $\theta_1 = \theta_3$, il vient

$$2\,\theta_1 + 24 \times 156 = 5000\,\mu\text{s}$$

et

$$RC = \theta_1 = \theta_3 = 625\,\mu\text{s}$$

La constante de temps RC est matérialisée par une résistance ajustable de 10 kΩ et une capacité de 100 nF.

II. 2. 3. 3. Calcul de la largeur θ_f de l'impulsion S_f.

Nous avons vu que l'équation de la tension de sortie $-V(t)$ d'un intégrateur lors de la décharge de la capacité C dans R_f, s'écrivait, avec $R_f = R' + R_{0N}$:

$$- V(t) = - V(0). e^{-t/R_f. C}$$

où $-V(0)$ est le potentiel aux bornes de C avant la décharge.

La constante de temps $R_f. C$ est fixée a priori ; on l'a prise égale à 180 μs , c'est-à-dire, puisque C = 100 nF et $R_{0N} \# 50\ \Omega$, que cette valeur est ajustée en prenant $R' \# 1750\ \Omega$.

La durée θ_f de la décharge est telle que l'on ait

$$- V(\theta_f) = - \frac{3}{4} V(0)$$

Par suite

$$\frac{3}{4} = e^{- \theta_f/R_f. C}$$

$$\theta_f = R_f. C \text{ Log } \frac{4}{3} = 180. 0,287 \mu s$$

et finalement

$$\theta_f = 51,7 \mu s$$

II. 2. 3. 4. Calcul de la constante de temps des monostables 1, 2, 3, 4 et 5.

* Monostables 1 et 3.

On s'est imposé $\theta_1 = \theta_3 = 625\ \mu s$

aussi

$$R_1 C_1 = R_3 C_3 = \frac{\theta_1}{\text{Log } 2} = 893 \mu s$$

Les composants utilisés sont

$R_1 = R_3 = 22\ k\Omega$ ajustable

$C_1 = C_2 = 100$ nF

* Monostable 2 .

En simulation, l'opération d'intégration d'une impulsion sur les voies 25 , 26 et 27 est définie par

$$I_t = I'_{t-1} + 4$$

ce qui implique analogiquement, si l'on garde la même constante d'intégration que sur les autres voies

$$\theta_2 = \frac{\theta_1}{5} = 125 \mu s$$

Ainsi le circuit de temporisation du monostable 2 est caractérisé par

$$R_2 C_2 = \frac{125}{\text{Log } 2} = 179 \mu s$$

où

$R_2 = 22$ kΩ ajustable

$C_2 = 10$ nF

* Monostable 4 .

L'impulsion r_4 , en verrouillant la porte II, inhibe le déclenchement du monostable 2 lors des seize derniers décalages dans le registre R (voir planche 5) ; il faut donc

$$\theta_4 \geqslant 16 \times 156 \mu s$$

On choisit

$$\theta_4 = 2,8 \text{ ms}$$

par conséquent

$$R_4 C_4 = \frac{2,8}{\text{Log } 2} = 4 \text{ ms}$$

avec

$R_4 = 47$ kΩ ajustable

$C_4 = 100$ nF

* Monostable 5 .

Ce monostable élabore l'impulsion r_5 qui, composée dans la porte V avec r_3 , forme l'impulsion S_f de largeur θ_f , telle que

$$\theta_f = \theta_5 - \theta_3 = 51,7 \mu s$$

Ainsi

$$\theta_5 = 676{,}7\ \mu s \qquad \text{et} \qquad R_5 C_5 = \frac{676{,}7}{\text{Log}\,2} = 968\ \mu s$$

avec

$R_5 = 22\ k\Omega$ ajustable

$C_5 = 100\ nF$

II. 3. Calcul des résistances de la matrice d'apprentissage.

Nous ne reviendrons pas sur la fonction de la matrice d'apprentissage dans le système de reconnaissance, ni sur sa structure (on se référera au paragraphe I. 2. 3. pour la matrice simulée et au paragraphe II. 1. pour la matrice réalisée) ; rappelons simplement que sur la $i^{ième}$ ligne, la grandeur de sortie SM_i, calculée par simulation, s'écrit :

$$SM_i = K \,.\, P_{1i} + \sum_{j=2}^{30} P_{ij} \,.\, EM_j \qquad (5)$$

où les P_{ij} sont les poids de la matrice et les EM_j (avec $EM_1 = K$) les grandeurs d'entrée de celle-ci, comprises entre - 25 et + 25 (voir planche 8).

Le dispositif choisi pour élaborer ces sorties SM_i sous forme analogique est une matrice de résistances (cf. planche 4) dont les entrées sont les tensions E_j, dédoublées en E_j^+ et E_j^- afin de matérialiser le signe des poids calculés (avec en particulier $E_1^+ = K'$ et $E_1^- = -K'$), et les sorties, les potentiels S_i produits aux bornes des résistances de charge R_L.

L'impédance interne des générateurs de tension connectés sur les entrées de la matrice de résistances étant très faible, nous pouvons considérer que toutes les lignes de cette matrice sont découplées entre elles et, par conséquent, calculer les résistances relatives à une ligne, indépendamment de celles relatives aux autres lignes.

Considérons le schéma électrique associé à la $i^{ième}$ ligne (figure 6).

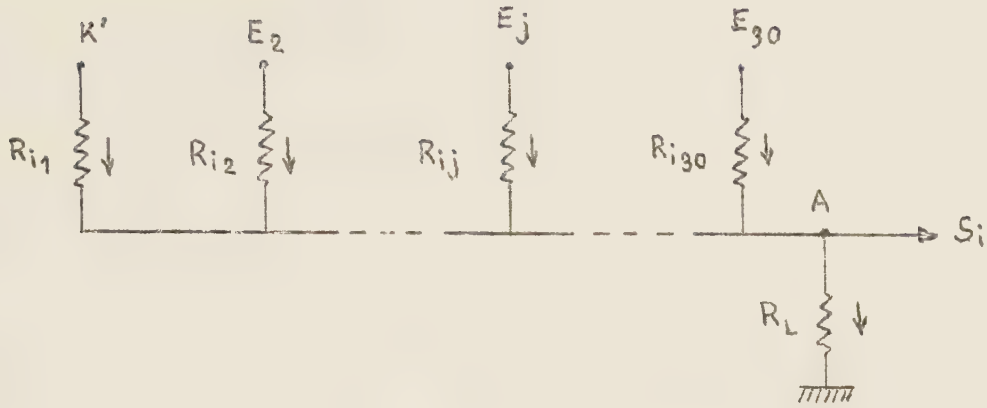


Figure 6.

Si l'on convient du sens des courants donné sur la figure 6, on a au point A

$$\frac{K' - S_i}{R_{i1}} + \sum_{j=2}^{30} \frac{E_j - S_i}{R_{ij}} = \frac{S_i}{R_L}$$

ou encore

$$\frac{K'}{R_{i1}} + \sum_{j=2}^{30} \frac{E_j}{R_{ij}} = S_i \left[\sum_{j=1}^{30} \frac{1}{R_{ij}} + \frac{1}{R_L} \right]$$

Avec

$$G_i = \sum_{j=1}^{30} \frac{1}{R_{ij}} + \frac{1}{R_L}$$, conductance de la ligne i ,

la tension de sortie S_i a pour expression

$$S_i = \frac{K'}{G_i R_{i1}} + \sum_{j=2}^{30} \frac{E_j}{G_i R_{ij}} \qquad (6)$$

Les équations (5) et (6) apparaissent alors sous deux formes identiques à un coefficient ℓ près. Ce coefficient est obtenu en formant le rapport de la valeur maximale des SM_i calculés et de la tension maximale S_i autorisée sur les sorties de la matrice de résistances, cette limite étant imposée par le convertisseur analogique - numérique utilisé dans l'identificateur de maximum.

En se référant aux résultats rassemblés planche 8, on détermine

$$\ell = \frac{SM_i \max}{S_i \max} = \frac{3.10^5}{5} = 6.10^4$$

Ainsi, toute grandeur de sortie SM_i est liée à S_i par la relation

$$SM_i = \ell . S_i \qquad \text{avec} \qquad \ell = 6.10^4$$

Rappelons que pour les grandeurs d'entrée EM_j et E_j de la matrice, respectivement simulée et réalisée, nous avions :

$$EM_j = \ell_1 . E_j \qquad \text{avec} \qquad \ell_1 = 5 \quad , \quad EM_1 = K \quad \text{et} \quad E_1 = K'$$

Par conséquent, en remplaçant SM_i , K et EM_j par leur valeur dans l'équation (5), nous obtenons :

$$\ell . S_i = \ell_1 . K' . P_{1i} + \sum_{j=2}^{30} P_{ij} . \ell_1 . E_j$$

et

$$S_i = K' . \frac{\ell_1}{\ell} . P_{1i} + \sum_{j=2}^{30} \frac{\ell_1}{\ell} . P_{ij} . E_j \qquad (7)$$

L'identification des deux expressions (6) et (7), de la tension de sortie S_i générée sur la $i^{ième}$ ligne, conduit à écrire

$$\frac{\ell_1}{\ell} . |P_{ij}| = \frac{1}{G_i . R_{ij}}$$

relation dans laquelle les poids P_{ij} sont pris en valeur absolue puisque leur signe est traduit par E_j dédoublé.

En d'autres termes, avec $\ell_2 = \frac{\ell}{\ell_1} = 12000$, et les poids réajustés $p_{ij} = \frac{P_{ij}}{\ell_2}$ on a ,

$$p_{ij} = \frac{1}{G_i . R_{ij}} \qquad (8)$$

Cherchons à exprimer la conductance d'une ligne G_i en fonction des poids réajustés p_{ij}.

Par définition

$$G_i = \sum_{j=1}^{30} \frac{1}{R_{ij}} + G_L$$

où $G_L = \frac{1}{R_L}$ est la conductance de charge.

Ainsi, avec (8), on a

$$G_i = \sum_{j=1}^{30} G_i . p_{ij} + G_L$$

donc

$$G_i = \frac{G_L}{1 - \sum_{j=1}^{30} p_{ij}}$$

avec comme condition d'existence $\sum_{j=1}^{30} p_{ij} < 1$.

En remplaçant l'expression de G_i dans (8), il vient alors

$$R_{ij} = \frac{1 - \sum_{j=1}^{30} p_{ij}}{G_L . p_{ij}} \qquad (9)$$

Le calcul des résistances R_{ij} à partir des poids calculés P_{ij} se résume donc à :

- transposer les P_{ij} en p_{ij} ;

- vérifier la condition $\sum_{j=1}^{30} p_{ij} < 1$;

- se fixer la conductance de charge G_L qui est la même pour toutes les lignes de la matrice ;

- calculer les conductances de ligne $G_i = \frac{G_L}{1 - \sum_{j=1}^{30} p_{ij}}$;

- puis, former pour chaque résistance l'expression

$$R_{ij} = \frac{1}{G_i . p_{ij}}$$

Ce calcul a été effectué sur ordinateur.

Remarque : Les courants débités par les sources de tension E_j étant fonction des résistances R_{ij}, il est intéressant, afin de limiter ces courants, de se fixer, en plus de la condition $\sum_{j=1}^{30} p_{ij} < 1$, une valeur minimale pour R_{ij} de $500\,\Omega$.

Ainsi, d'après la relation (9), la différence $1 - \sum_{j=1}^{30} p_{ij}$ doit, au moins, être 500 fois supérieure au produit $G_L \cdot p_{ij}$. Si cette contrainte n'est pas respectée, on procède alors à un nouveau réajustement des poids P_{ij} par un coefficient $\ell_3 > \ell_2$. L'opération se répète tant que les deux conditions $\sum_{j=1}^{30} p_{ij} < 1$ et $R_{ij} \geqslant 500\,\Omega$ ne sont pas satisfaites simultanément.

La revalorisation de ℓ_2, si elle a lieu, a pour effet de diminuer la plage de variation des tensions de sortie S_i que l'on avait initialement fixée à ± 5 V. Il faudra naturellement veiller dans le calcul à ce que cette plage reste suffisamment grande (± 4 V minimum) afin de conserver entre les tensions S_i un écart compatible avec la résolution du convertisseur analogique - numérique de l'identificateur de maximum.

Les résistances utilisées dans la réalisation de la matrice d'apprentissage ont été choisies avec une précision inférieure à 1 % par rapport à la valeur calculée. Cette précision a également été obtenue sur les sorties de la matrice lors du test systématique de toutes les lignes.

II. 4. Identificateur de maximum.

Cette partie du dispositif de reconnaissance réalise l'identification de la sortie de la matrice d'apprentissage présentant le potentiel le plus élevé par rapport à celui des autres sorties. Le numéro de la ligne détectée correspondra au numéro du phonème reconnu.

II. 4. 1. Fonctionnement général et synchronisation d'un cycle de tests.

Le schéma synoptique de l'identificateur est donné planche 9 ; les diagrammes des temps des signaux qui y sont élaborés, planche 10, figure a.

Trois fonctions principales définissent cet ensemble :

- le multiplexage analogique des 30 voies d'entrée, c'est-à-dire la mise en correspondance successive de chacune des voies, avec la sortie unique E_C d'un multiplexeur ;

- la conversion analogique - numérique de la tension E_C ;

- la comparaison digitale de deux nombres codés en binaire pur.

On convient de stocker :

- dans le registre I, l'équivalent numérique RI de la tension présente sur la voie sélectionnée ;

- dans le compteur III, le numéro de cette voie ;

- dans le registre II, l'équivalent numérique RII de la plus grande des tensions jusqu'alors converties ;

- dans le registre III, le numéro de la voie portée au "potentiel" digital RII.

Les sorties du registre I et du compteur III constituent respectivement les entrées du registre II et du registre III.

Le départ d'un cycle de tests est donné par l'impulsion de calibrage des fuites S_f. Cette impulsion remet simultanément à zéro le compteur III, l'adressage du multiplexeur et le convertisseur analogique-numérique, commutant ainsi la première voie vers l'entrée E_C du convertisseur. Le front de montée de S_f coïncide avec le début de la conversion et provoque le basculement du signal STATUS, d'un niveau bas à un niveau haut. Ce signal de service est élaboré par le convertisseur A/N et signifie, quand il est à l'état logique 1 , qu'une conversion est en cours.

Lorsque celle-ci se termine, STATUS retombe au niveau O et déclenche, après passage dans la porte I, le monostable 6. L'impulsion r_6 générée valide le transfert en parallèle du résultat de la conversion dans le registre I (c'est le signal HRI), et incrémente sur son front de descente le compteur d'adressage du multiplexeur (signal H_M). La voie suivante est alors sélectionnée.

La comparaison des nombres binaires RI et RII contenus respectivement dans les registres I et II s'effectue en parallèle :

- si RI > RII , la sortie C_D du comparateur digital est à l'état 1, forçant la sortie de la porte III dans le même état, puisqu'il s'agit d'un opérateur OU.

- si RI ⩽ RII , $C_D = 0$.

Le résultat de la comparaison est pris en compte dans la porte IV (opérateur ET) par l'impulsion r_7 formée dans le monostable 7 , après déclenchement de celui-ci sur le front descendant de r_6.

En effet, l'équation du signal de sortie de cette porte est :

$$H\,R\,II = H\,R\,III = r_7\,(DC\,1 + C_D)$$

Si nous supposons dans un premier temps que DC 1 = 0, l'équation devient $H\,R\,II = H\,R\,III = r_7 . C_D$; elle exprime que C_D valide (pour $C_D = 1$) ou inhibe (pour $C_D = 0$) l'impulsion r_7 sur les entrées de prédétermination H R II et H R III.

Par conséquent :

- si R I > R II , $C_D = 1$ et $H\,R\,II = r_7$ ordonne le transfert parallèle de R I dans le registre II, $H\,R\,III = r_7$ ordonne celui du contenu du compteur III dans le registre III ;

- si R I ⩽ R II , $C_D = H\,R\,II = H\,R\,III = 0$ et la configuration des registres II et III reste inchangée.

L'impulsion r_7 (composée avec S_f dans l'opérateur OU de la porte II) réinitialise également le convertisseur A/N et commande le départ de la conversion sur la voie qui vient d'être multiplexée. Un nouveau test commence alors.

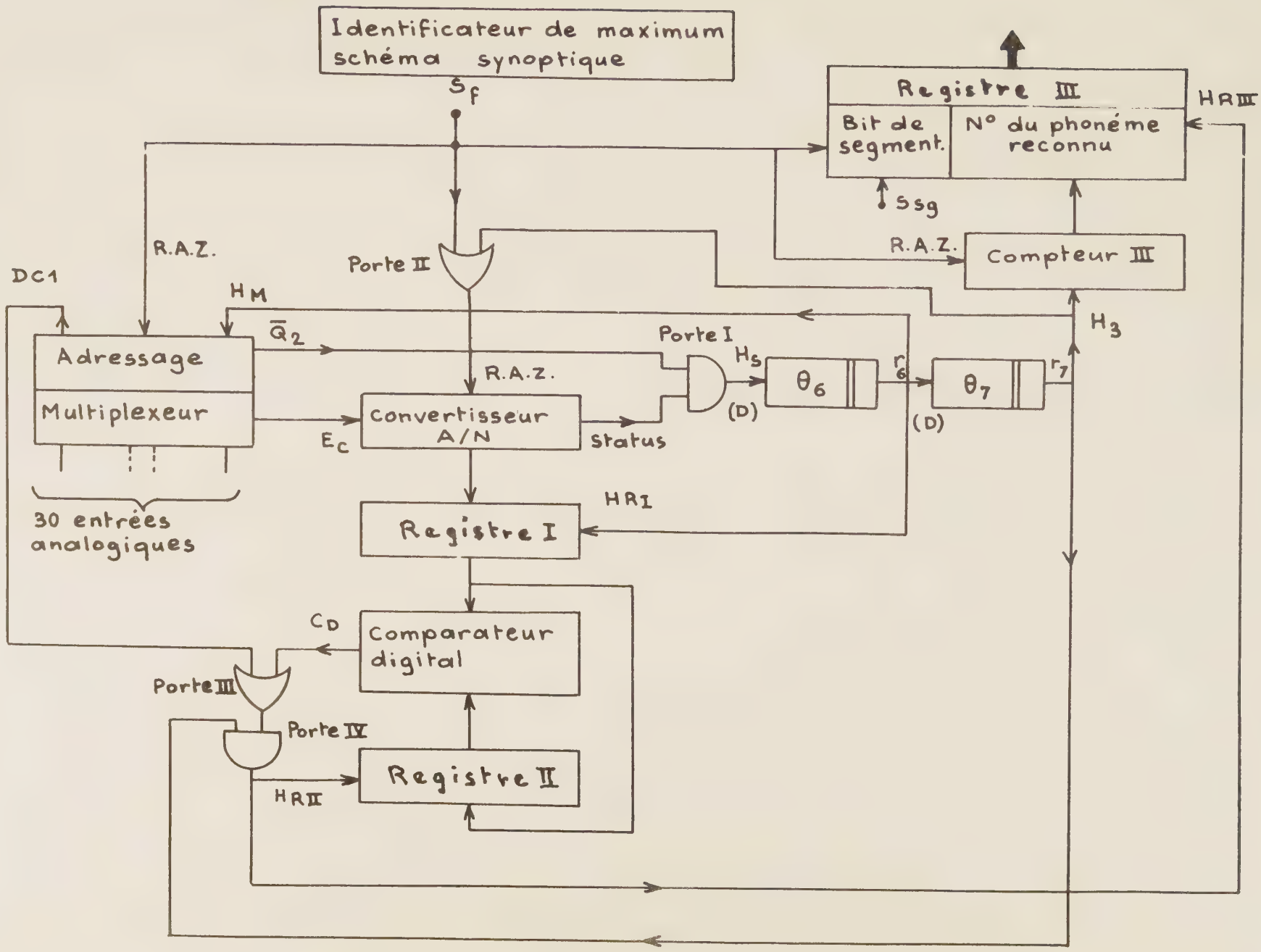
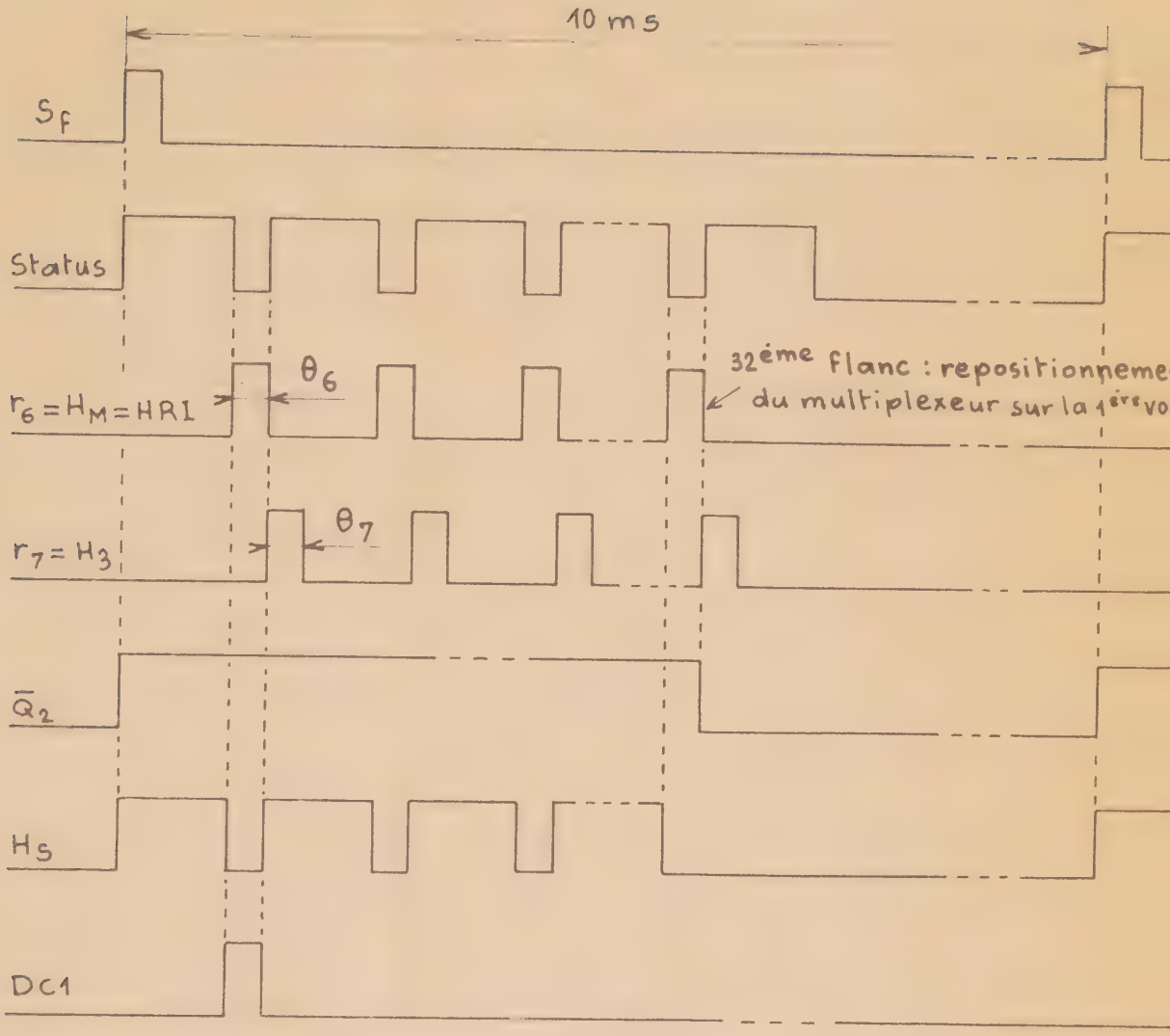

Identificateur de maximum schéma synoptique
S_f
Registre III
Bit de segment.
N° du phonéme reconnu
HRIII
S_{sg}
R.A.Z.
Compteur III
R.A.Z.
Porte II
DC1
H_M
$\overline{Q}_2$
Adressage
Multiplexeur
30 entrées analogiques
Porte I
H_s
θ_6
r_6
θ_7
r_7
H_3
(D)
(D)
R.A.Z.
Convertisseur A/N
E_c
Status
HRI
Registre I
C_D
Comparateur digital
Porte III
Porte IV
Registre II
HRII

Planche 9



$$\begin{cases} \theta_6 = 10\,\mu s \\ \theta_7 = 10\,\mu s \end{cases}$$

fig a : Diagramme des principaux signaux générés lors de l'identification du maximum

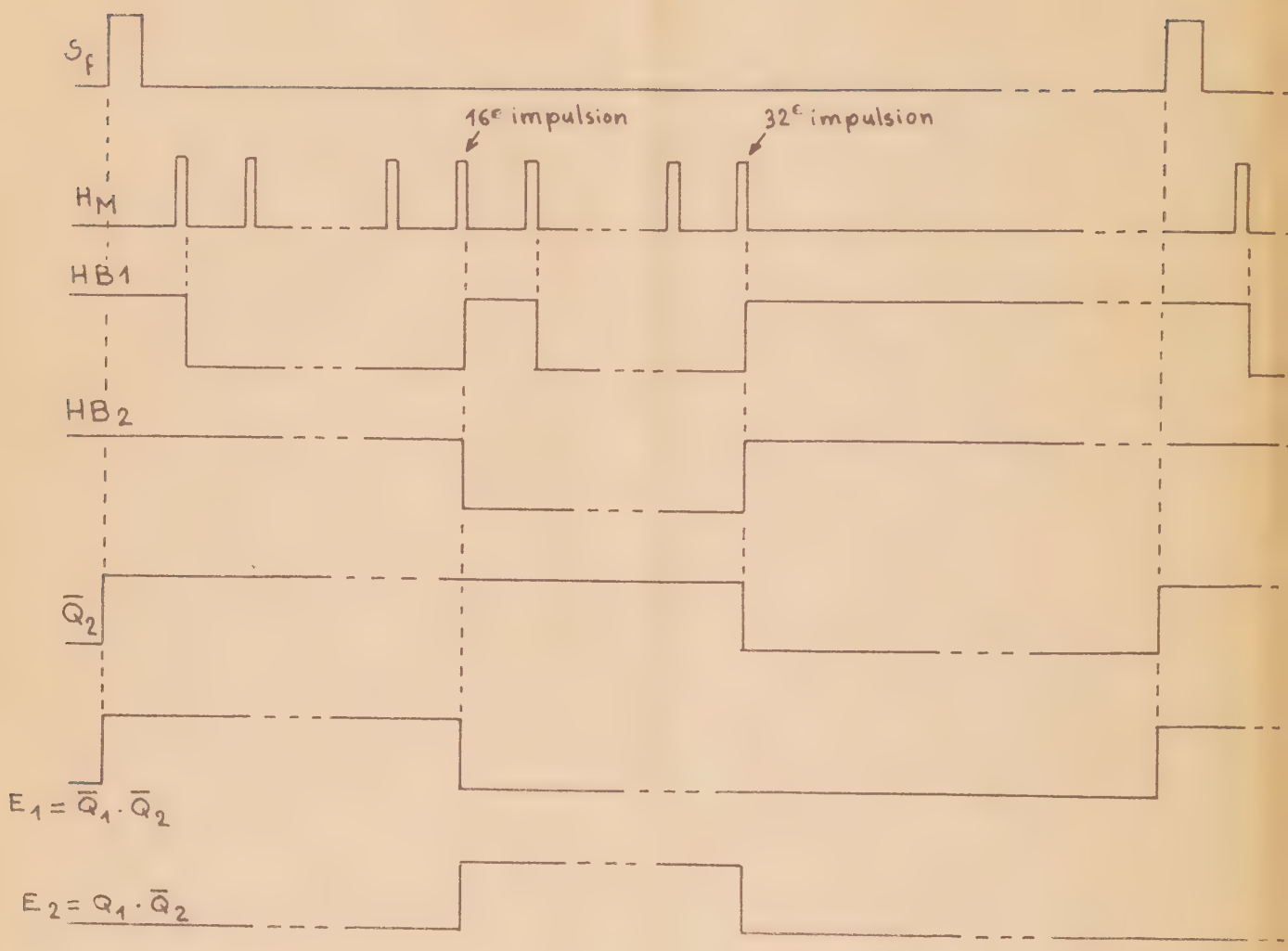


fig b : Diagramme des signaux d'adressage du multiplexeur

Les 30 voies d'entrée sont ainsi traitées de façon identique, avec cependant une particularité pour le test de la première : le registre II n'étant jamais remis à zéro, on est obligé, dans ce cas, de forcer le résultat R I de la première conversion dans ce registre quelle que soit la valeur de C_D ; c'est le signal DC 1 repérant la sélection de la première voie qui permet ce transfert puisqu'alors DC 1 = 1 , et par suite H R II = H R III = r_7.

La fin d'un cycle de tests coïncide avec le basculement du signal $\overline{Q_2}$ du niveau logique 1 au niveau 0 ; le verrouillage de la porte I, qui s'en suit, inhibe le déclenchement du monostable 6 et par conséquent l'adressage du multiplexeur.

Nous avons adjoint au numéro du phonème contenu dans le registre III de sortie, un bit de segmentation S_{sg} dont nous préciserons le mode d'élaboration dans la dernière partie de ce travail.

II. 4. 2. Fonctionnement du multiplexeur.

Le multiplexeur utilisé comprend deux boîtiers (réf : DG 506 BR Siliconix) ; chacun des boîtiers assure le multiplexage de 16 voies dont les potentiels peuvent varier de - 10 à + 10 V. Le brochage interne d'un boîtier est donné figure 7.

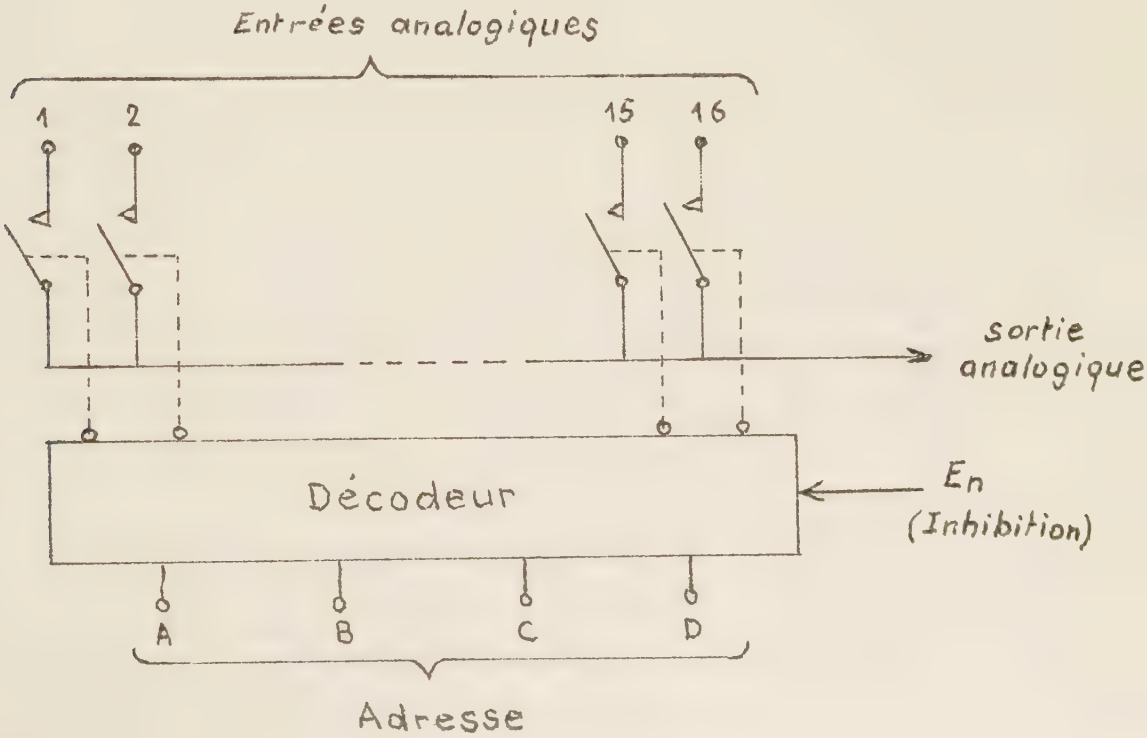


Figure 7.

La sélection d'une voie se fait par adressage d'un mot codé en binaire pur, sur les entrées A, B, C, D (A étant le bit de poids le plus fort). Cette adresse est décodée et valide la fermeture de l'interrupteur analogique correspondant (il s'agit de transistors MOS complémentaires).

Ainsi, une seule entrée est commutée sur la sortie du boîtier. On dispose également d'une entrée d'inhibition E_n qui, lorsqu'elle est au niveau logique O, annule l'adressage quel qu'il soit.

Le schéma synoptique du multiplexage est donné planche 11, le diagramme des temps des signaux planche 10, figure b.

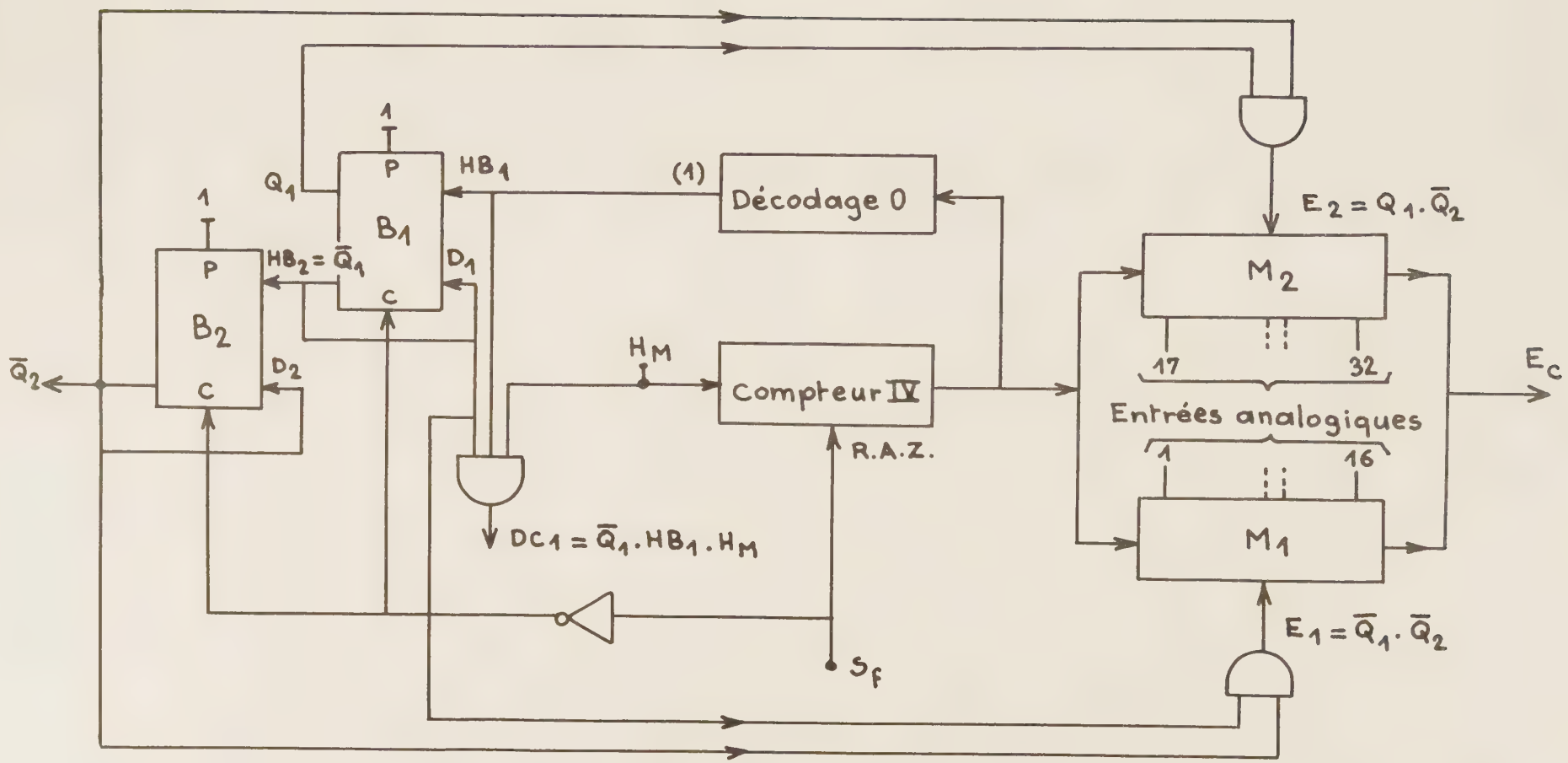
Les deux boîtiers multiplexeurs M1 et M2 sont commandés par un même mot binaire de quatre bits pris en parallèle sur les sorties du compteur IV, l'adressage n'étant validé que sur un seul des boîtiers. Leurs sorties analogiques sont reliées et forment l'entrée E_C du convertisseur A/N.

En début de multiplexage, le signal S_f réinitialise le compteur IV à zéro et force à l'état 0 les sorties Q_1 et Q_2 des bascules B 1 et B2.

Ces bascules, de type D (circuit Texas Instruments SN 7474 N), transfèrent l'information d'entrée D sur la sortie Q, lors du front montant de l'impulsion d'horloge HB . P et C sont des entrées à niveau permettant de mettre la sortie dans l'état logique 0 ou 1 quel que soit l'état de l'horloge.

Les bascules sont ici montées en cascade, la sortie $\overline{Q}$ de chacune d'elles est réinjectée à l'entrée. Le signal d'horloge HB1 de la première, qui correspond au décodage de la position 0 du compteur IV, est ainsi divisé en fréquence : par deux sur la sortie Q_1 et par quatre sur la sortie Q_2.

Lorsque S_f est appliqué, le produit $E_1 = \overline{Q}_1 . \overline{Q}_2$ vaut 1, tandis que le produit $E_2 = Q_1 . \overline{Q}_2$ vaut 0 ; par conséquent, seul le multiplexage des seize entrées du boîtier M1 est validé. La commutation s'effectue successivement après chaque incrémentation du compteur d'adressage sur le flanc descendant de l'impulsion $H_M = r_6$.



Adressage du multiplexeur

Planche 11

A la seizième impulsion H_M, le compteur IV est repositionné à zéro, la bascule B1 déclenchée et les entrées d'inhibition E_1 et E_2 portées respectivement au niveau 0 et 1. Le multiplexage des seize entrées de M 2 s'opère alors de la même façon que ci-dessus (notons que les entrées 31 et 32 sont reliées à la masse, la matrice d'apprengissage possédant 30 sorties).

La scrutation de toutes les voies se termine sur le flanc descendant de la 32ème impulsion H_M. A cet instant, le repositionnement du compteur à zéro provoque la déclenchement des bascules B1 puis B2 et l'apparition d'un niveau 0 sur la sortie $\overline{Q}_2$. ce niveau est utilisé pour verrouiller le signal H_S dans la porte I (cf. paragraphe II. 4. 1).

La logique d'adressage ci-dessus est complétée par une porte dans laquelle est formé le signal DC1 qui, lors de la sélection de la première voie, synchronise la prédétermination des registres II et III quelle que soit la réponse du comparateur digital. Il a pour équation

$$DC1 = HB1 \cdot \overline{Q}_1 \cdot H_M$$

Au cours d'un cycle de tests, H B1 repère les deux passages du compteur IV par zéro (sélection des première et septième voies), $\overline{Q}_1$ associé à H B1 valide le repérage du premier passage (première voie) et H_M permet d'exprimer le résultat du produit $HB1 . \overline{Q}_1$ sous forme impulsionnelle.

II. 4. 3. Convertisseur analogique - numérique. [11]

Le convertisseur utilisé (module ADC 8 H Analog Devices), ainsi que le diagramme des signaux générés est représenté planche 12.

La conversion analogique - numérique du signal d'entrée s'effectue avec une résolution de 8 bits. La tension pleine échelle du convertisseur est de 10 V ; elle définit l'intervalle de variation des tensions d'entrée - ici, de - 5 V à + 5 V -.

On associe :

- à la borne inférieure - 5 V le nombre binaire 000....0,

- à la borne supérieure + 5 V le nombre binaire 111....1;

en particulier, une tension nulle se traduit numériquement par le nombre 100....0. Remarquons également que le bit de poids le plus fort (MSB) peut être utilisé comme bit de signe : il vaut 0 pour une tension d'entrée négative et 1 pour une tension positive.

Deux parties constituent l'ensemble du convertisseur :

- une partie logique, qui élabore des informations numériques sur les sorties de huit bascules, par l'intermédiaire d'une horloge interne et d'un compteur de séquences,

- une partie analogique, dans laquelle cette information est restituée analogiquement par sommation de courants pondérés, puis comparée à la tension que l'on veut convertir.

La conversion se réalise en série, par approximations successives ; le déroulement des opérations est le suivant :

- les bascules et les compteurs sont initialisés à zéro ;

- lorsqu'une impulsion d'horloge est appliquée, une des huit bascules se trouve forcée dans l'état 1 (la commutation se fait du poids le plus fort (MSB) vers le poids le plus faible (LSB));

- l'information binaire ainsi formée sur les sorties des bascules est convertie en tension, puis comparée à la tension d'entrée ;

- le comparateur ordonne le maintien ou le refus du poids commuté ; cette décision étant prise sur le front montant d'une impulsion d'horloge.

Ainsi, à la fin de la huitième pesée, l'équivalent numérique de la tension convertie est disponible en parallèle sur les sorties des bascules. La disponibilité de cette information est plus précisément repérée par l'état du signal d'occupation STATUS qui passe au niveau logique 1 dès le départ d'une conversion, puis reste à ce niveau jusqu'à la fin de celle-ci.

Le diagramme donné planche 12 figure b montre, à titre indicatif, comment se construit sur les sorties, la tension numérique codée 00110110.

Fig a Schéma du convertisseur

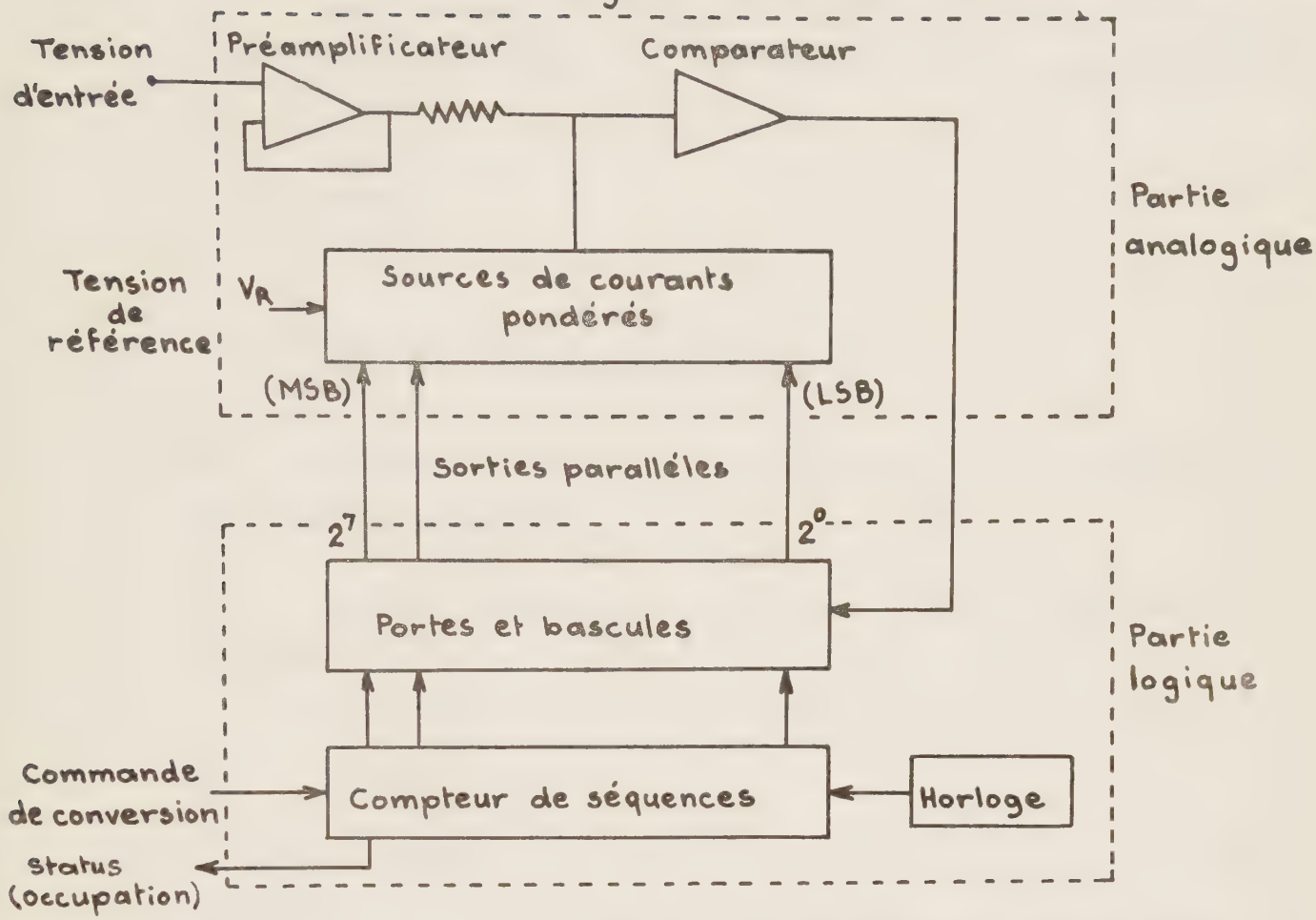
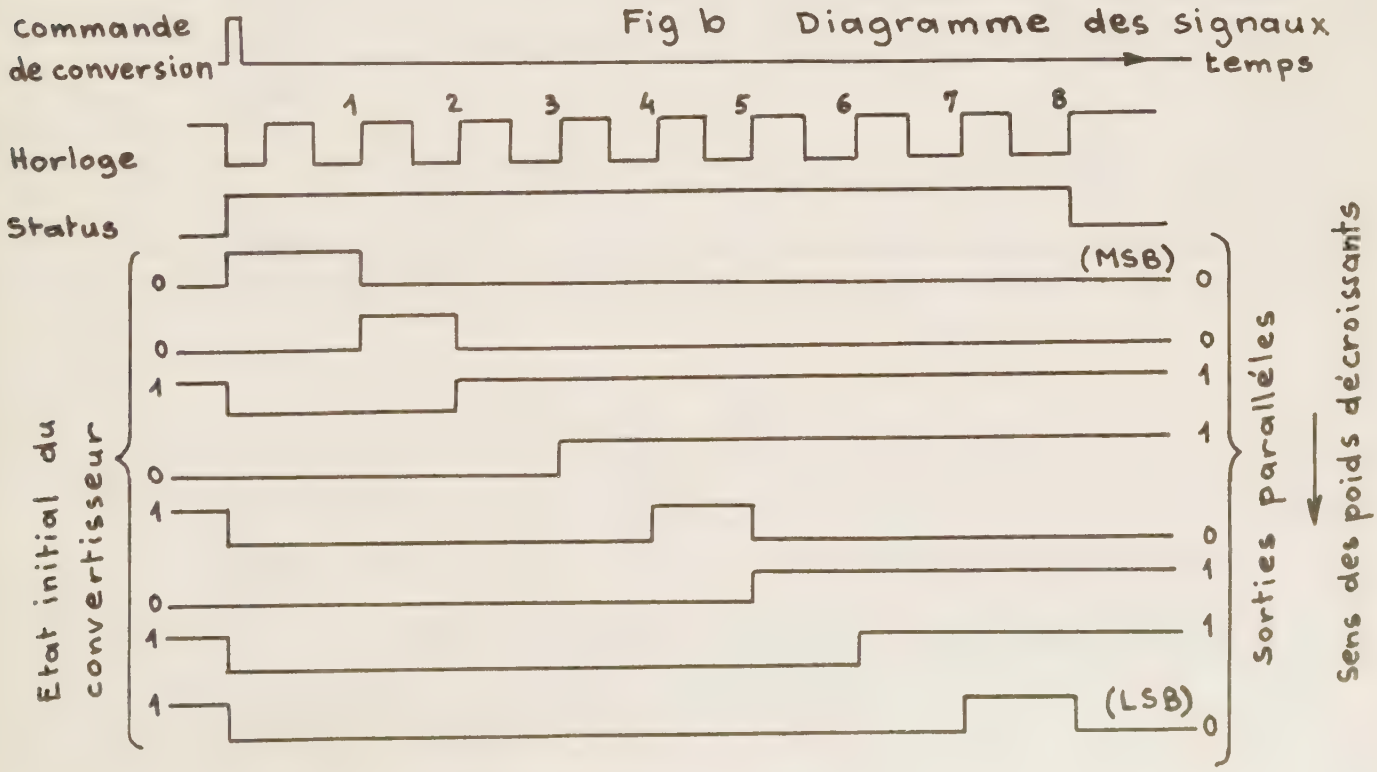


Fig b Diagramme des signaux



II. 4. 4. Comparateur digital.

II. 4. 4. 1. Principe de la comparaison.

Les nombres RI et R II stockés respectivement dans le registre I et II sont codés en binaire pur et définis par 8 bits. On veut réaliser le test RI > RII .

Soient

$$RI = A_p = a_p\, a_{p-1} \ldots\ldots a_i \ldots\ldots a_2 a_1$$

$$RII = B_p = b_p\, b_{p-1} \ldots\ldots b_i \ldots\ldots b_2 b_1$$

où a_i représente un bit de A_p , et b_i un bit de B_p , p étant le rang du bit de poids le plus élevé.

Vérifier l'inéquation $(A_i > B_i)$, c'est vérifier la relation de récurrence,

$$(A_i > B_i) = (a_i > b_i) + (a_i = b_i)(A_{i-1} > B_{i-1})$$

avec $A_1 > B_1 = (a_1 > b_1)$

qui exprime en effet que le nombre binaire A_i est strictement supérieur au nombre B_i quand le bit de poids le plus fort a_i est strictement supérieur à b_i , ou dans le cas d'une égalité, quand le test $(A_{i-1} > B_{i-1})$ est positif.

On a schématisé planche 1 3, figure a, la cellule de comparaison correspondant à cette relation et dans laquelle interviennent les opérateurs booléens somme, produit, supérieur et égal.

En remplaçant $(A_{i-1} > B_{i-1})$ puis $(A_{i-2} > B_{i-2})$ jusqu'à $(A_1 > B_1)$ par leur cellule équivalente, on peut représenter l'équation de récurrence par une juxtaposition de cellules de comparaison qui tiennent compte de la réponse élaborée dans la cellule d'ordre inférieur (cf. planche 1 4).

Les tables de vérité des fonctions logiques "supérieur" et "égal" établies sur deux variables binaires sont données ci-après :

a_i	b_i	$a_i = b_i$
0	0	1
0	1	0
1	0	0
1	1	1

a_i	b_i	$a_i > b_i$
0	0	0
0	1	0
1	0	1
1	1	0

On en tire les équations $(a_i = b_i) = \bar{a}_i \bar{b}_i + a_i b_i$

$$(a_i > b_i) = a_i \bar{b}_i$$

L'équation récurrente s'écrit alors

$$(A_i > B_i) = a_i \bar{b}_i + (\bar{a}_i \bar{b}_i + a_i b_i)(A_{i-1} > B_{i-1})$$

avec $A_1 > B_i = a_1 \bar{b}_1$

II.4.4.2. Réalisation du comparateur.

La réalisation d'une cellule de comparaison nécessite l'utilisation de deux double opérateurs ET - OU - NON (circuit Texas Instruments SN 7451 N) et d'un inverseur (cf. planche 13, figure b).

L'ensemble du comparateur digital est schématisé planche 14 où une cellule de comparaison est affectée au bit de même poids de chaque registre. Les registres I et II sont formés de huit bascules du type D à sorties complémentaires Q et $\bar{Q}$ (circuit Texas SN 7475N). La sortie Q d'une bascule suit l'information présente en D, tant que l'horloge est au niveau haut ; quand celle-ci passe au niveau bas, Q reste dans le même état que D au moment de la transition.

La prédétermination des registres I et II est ordonnée respectivement par les signaux d'horloge $HRI = r_6$ et $HRII = (DC1 + C_D)r_7$. Lorsque $HRI = 1$, le mot binaire du convertisseur analogique - numérique est transféré en parallèle sur les sorties du registre I pour former le mot A_8 ; lorsque $HRII = 1$, le mot binaire A_8 est transféré en parallèle sur les sorties du registre II et constitue le mot B_8.

* * * * * * * *

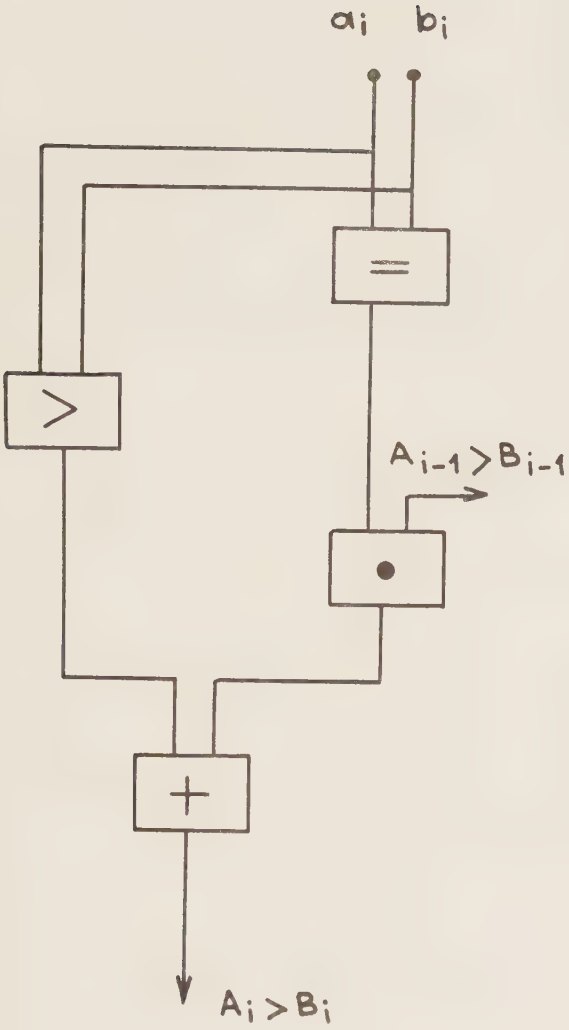


fig a

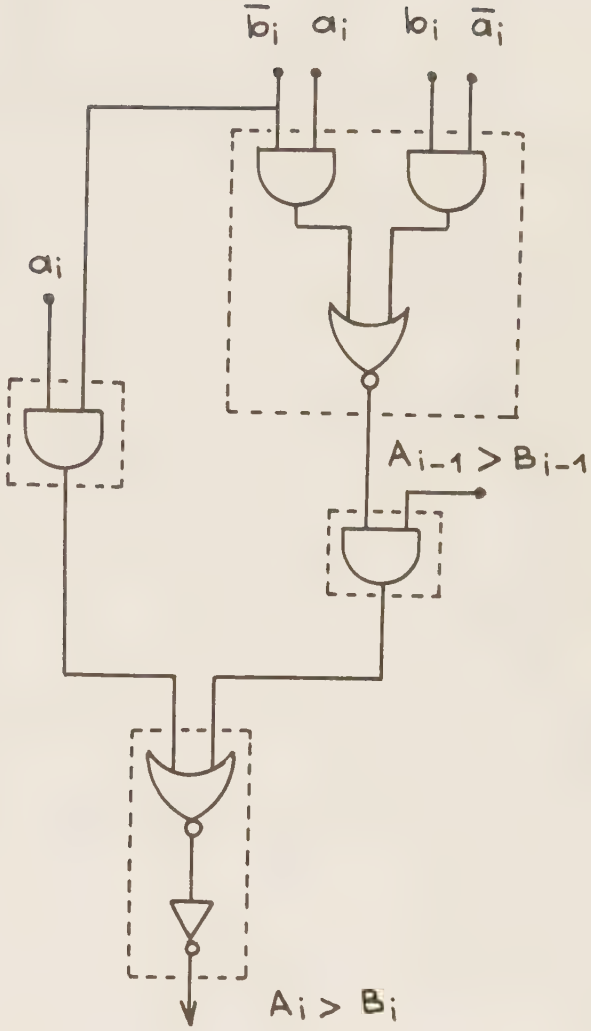
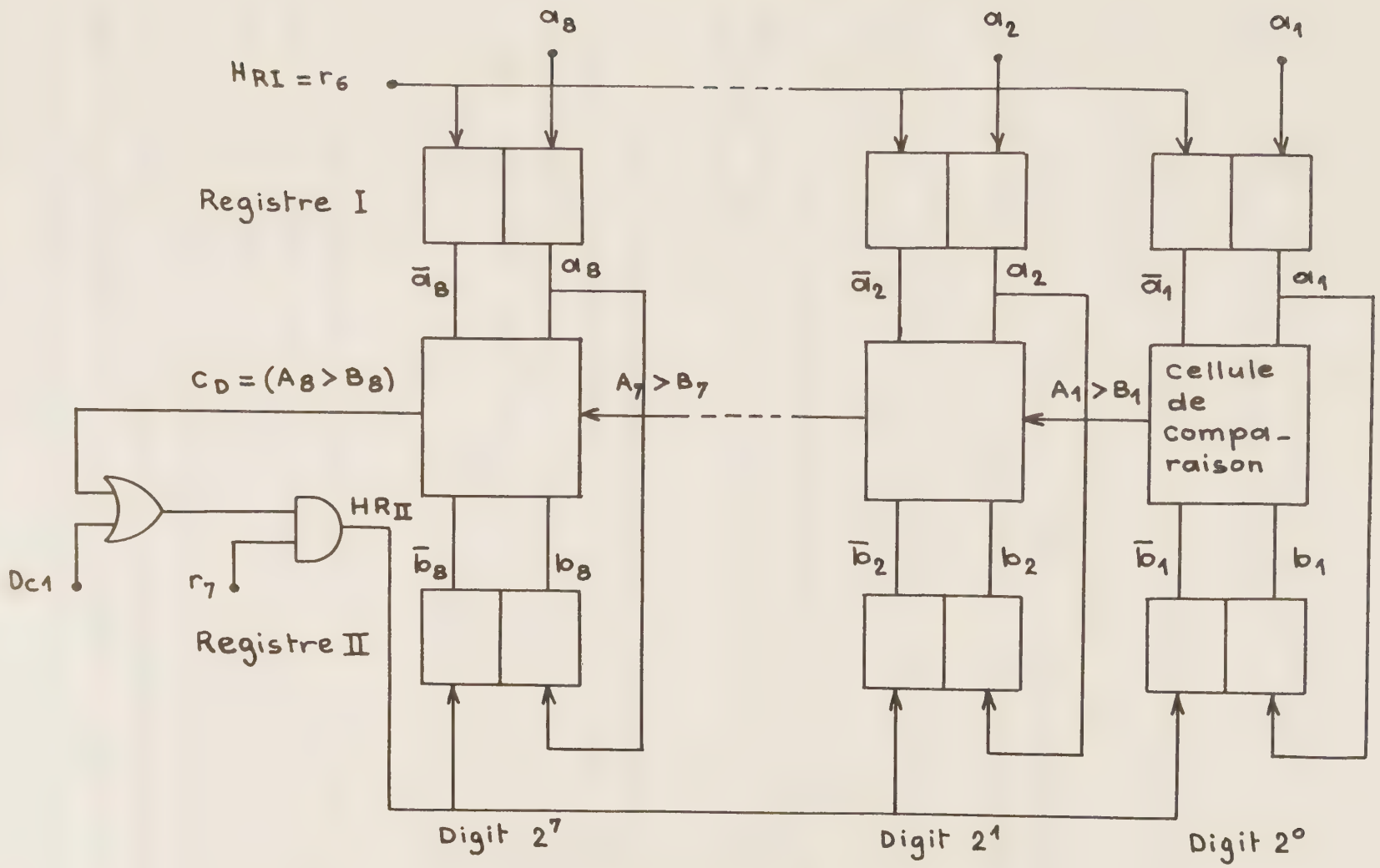


fig b

La fonction $A_i > B_i$ et sa réalisation à l'aide de deux double opérateurs OU inversé.

Planche 13



Comparateur digital

Planche 14

CHAPITRE III - REALISATION D'UN SYSTEME DE SEGMENTATION PHONEMIQUE DE LA PAROLE.

Dans cette étude, la reconnaissance de la parole consiste à identifier les différents phonèmes qui composent un message parlé.

Du point de vue spectral, les phonèmes se divisent approximativement en deux catégories :

- ceux dont le spectre en fréquence reste sensiblement constant lorsque le temps évolue, donc caractérisés par un régime soutenu ;

- ceux dont le spectre en fréquence varie en fonction du temps et qui se traduisent par un régime transitoire.

Le signal de parole apparaît comme une succession de tels régimes, le passage entre deux régimes soutenus se faisant nécessairement par un régime transitoire.

Par conséquent, comme la fréquence d'échantillonnage en temps du signal vocal est plus grande que la fréquence d'apparition des phonèmes, on transférera dans le registre de sortie de l'appareil de reconnaissance, la même information pendant plusieurs échantillonnages, lorsque le phonème est soutenu, et une suite d'informations différentes s'il s'agit d'un phonème transitoire.

Nous abordons ici le délicat problème de la séparation des phonèmes ou segmentation phonémique en temps réel, d'un message parlé. En effet, si la détection du début et de la fin d'un phonème soutenu est relativement aisée - il suffit de repérer l'implantation d'une nouvelle valeur dans le registre de sortie -, il n'en est pas de même pour un phonème transitoire, car durant celui-ci les informations recueillies varient constamment et une variation ne coïncide pas forcément avec la fin du phonème.

La segmentation de la parole, si elle était rigoureuse, permettrait de valider sur les sorties du système de reconnaissance, un nombre d'informations correspondant exactement au nombre de phonèmes composant le message.

Nous exposons dans cette dernière partie le principe puis la réalisation d'un dispositif qui effectue la segmentation du signal vocal parallèlement à sa reconnaissance. Le résultat de ces opérations est inscrit dans un même registre de sortie (le registre III, cf. planche 9).

III.1. Principe de la segmentation.

Nous l'avons signalé, la segmentation peut se faire au niveau du registre de sortie de l'appareil de reconnaissance en observant les variations de l'information qu'il contient et qui sont la suite des numéros d'identification de chaque phonème. En fait, il est préférable d'utiliser comme données, celles fournies par le prétraitement, beaucoup plus spécifiques de l'aspect du signal vocal. De plus, nous le verrons par la suite, il devient possible d'évaluer le "taux de variation" de ces informations lors de deux échantillonnages successifs, donc d'établir avec plus de souplesse la distinction entre régime soutenu et régime transitoire.

On définit l'ensemble X_t dont les éléments sont $X_t(i)$, comme l'ensemble des 24 bits du mot élaboré par le prétraitement numérique à l'instant t d'échantillonnage ; soit $X_t = \left\{ X_t(i) \right\}$, i variant de 1 à 24.

Sur cet ensemble, construisons les grandeurs :

* $N_t = \sum_{i=1}^{24} X_t(i)$, nombre de bits à 1 de ce mot binaire, à l'instant t ,

* $n_t = \sum_{i=1}^{24} n_t(i)$, nombre de transitions calculé sur deux mots pris, l'un à l'instant t , l'autre à l'instant $t - 1$.

Nous appelons transition, le changement d'état (de 0 à 1 ou de 1 à 0) du bit i de X_t lors de deux échantillonnages successifs associé à la variable binaire $n_t(i)$ telle que

$$n_t(i) = 0 \quad \text{si} \quad X_{t-1}(i) = X_t(i)$$

$$n_t(i) = 1 \quad \text{si} \quad X_{t-1}(i) \neq X_t(i)$$

ou encore, en introduisant la fonction "OU exclusif" :

$$n_t(i) = X_{t-1}(i) \oplus X_t(i)$$

et par suite

$$n_t = \sum_{i=1}^{24} n_t(i) = \sum_{i=1}^{24} (X_{t-1}(i) \oplus X_t(i))$$

D'un point de vue formel

* $n_t = 0$ (donc $X_{t-1} = X_t$) caractérise un régime soutenu durant deux pas successifs d'échantillonnage,

* $n_t \neq 0$ ($X_{t-1} \neq X_t$) caractérise un régime transitoire.

En réalité les deux régimes ne se différencient pas aussi catégoriquement, car un régime soutenu ne se traduit pas physiquement par une stabilité rigoureuse des informations élaborées par le prétraitement. Il faut donc s'accorder une marge d'erreur sur n_t.

En outre, la simulation a établi que la discrimination entre les deux régimes se révèle beaucoup plus efficace si le nombre n_t est associé au nombre N_t de 1 présents dans un mot binaire à l'instant t et si les grandeurs n_t et N_t sont additionnées respectivement à n_{t-1} et N_{t-1} calculées à l'instant d'échantillonnage précédent. Ces opérations éliminent les fluctuations instantanées, tout en soulignant les transitoires réels de la parole.

Pour déterminer la nature du régime au temps t, nous formons alors

$$n_t^* = n_{t-1} + n_t$$

$$N_t^* = N_{t-1} + N_t$$

et nous comparons n_t^* à la somme $x + 0,5\ N_t^*$ (dans laquelle x peut s'assimiler à un seuil dépendant de la marge d'erreur que l'on s'est fixée) :

- si $n_t^* > x + 0,5\ N_t^*$, le régime est transitoire

- si $n_t^* \leq x + 0,5\ N_t^*$, le régime est soutenu.

Dans le plan d'abscisses N^*_t et d'ordonnées n^*_t (limité ici au premier quadrant puisque N^*_t et n^*_t sont tous deux positifs), résoudre ces inéquations revient à définir deux surfaces séparées par la droite d'équation $n^*_t = x + 0,5\ N^*_t$ (voir figure 8). L'ensemble des points $(N^*_t\ ,\ n^*_t)$ de la surface supérieure est relatif aux régimes transitoires, celui des points de la surface inférieure et de la droite de discrimination caractérise les régimes soutenus.

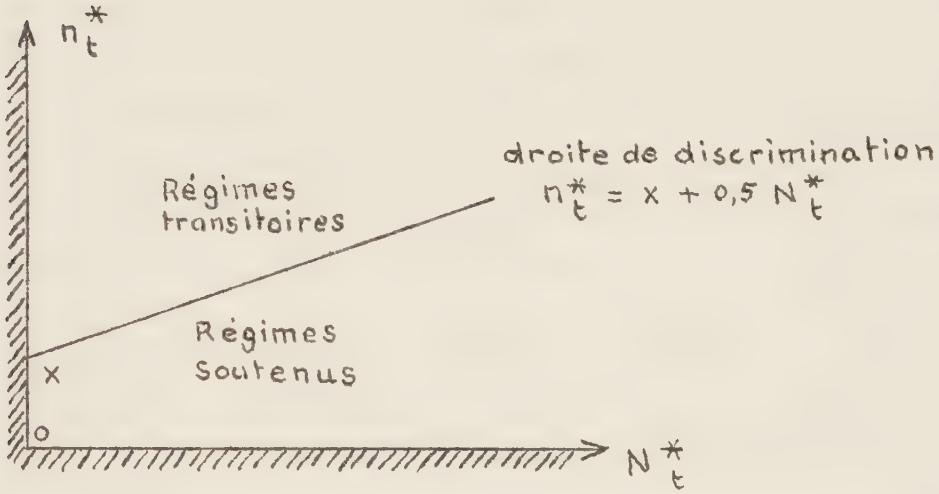


Figure 8.

L'expression de la loi de discrimination a été construite le plus simplement possible afin de réduire la complexité du dispositif de segmentation.

Nous réalisons par ces tests, à l'aide d'une seule variable binaire M_t, le codage du régime de parole à l'instant t. L'existence d'un régime transitoire T est associé à $M_t = 1$, celle d'un régime soutenu S à $M_t = 0$.

La segmentation consiste à repérer la transition d'un régime à un autre ; l'équation S_{sg} du bit de segmentation s'écrit par conséquent

$$S_{sg} = M_{t-1} \oplus M_t$$

La signification de cette information est la suivante :

- si $S_{sg} = 0$, $M_{t-1} = M_t$ et le régime de parole reste le même durant deux échantillonnages successifs, donc le phonème prononcé n'est pas terminé ;

- si $S_{sg} = 1$, $M_{t-1} \neq M_t$ et il y a transition (de S vers T ou de T vers S) ; en d'autres termes, un nouveau phonème débute au temps t .

La valeur S_{sg} est transférée toutes les 10^{-2} s dans le registre de sortie (registre III) de l'appareil de reconnaissance. Le synchronisme des opérations est conçu de telle sorte que S_{sg} élaboré au temps t soit en fait affecté au numéro du phonème reconnu au temps t-1 ; c'est-à-dire que le bit de segmentation, quand il vaut 1, repère la fin des phonèmes.

Cette remarque est illustrée sur la figure 9 où T symbolise un régime transitoire et S un régime soutenu.

	Phonème 1				Phonème 2			Phonème 3			Phonème 4			
Régimes	T	T	T	T	S	S	S	T	T	T	S	S	S	S
M_t	1	1	1	1	0	0	0	1	1	1	0	0	0	0
S_{sg}	0	0	0	1	0	0	1	0	0	1	0	0	0	1
Fin des phonèmes				↑			↑			↑				↑

temps →

Figure 9.

Notons que si cette façon de procéder est justifiée quant à la reconnaissance des phonèmes soutenus, elle est tout-à-fait arbitraire en ce qui concerne les phonèmes transitoires. En effet, de par sa définition, le phonème transitoire est difficilement identifiable à un instant précis, aussi, nous avons convenu de l'associer au numéro du phonème reconnu à la fin d'un régime transitoire. (Cette convention est celle choisie pour l'apprentissage simulé).

Les résultats sont néanmoins très satisfaisants : nous obtenons par simulation et pour un seuil x fixé à 1, un taux de segmentation de 140 %, c'est-à-dire que pour dix phonèmes composant un message, la segmentation repère quatorze zones différentes.

III. 2. Réalisation du système de segmentation.

Le schéma synoptique de l'ensemble et le diagramme des divers signaux sont donnés planche 15.

III. 2. 1. Fonctionnement général.

L'ensemble de segmentation comprend deux parties fonctionnant en parallèle. Chacune d'elles élabore un des membres de l'inéquation dont la résolution permet d'identifier le régime de parole à l'instant t et par suite, de repérer la transition entre les deux régimes possibles.

Le début des opérations coïncide avec le transfert, dans le registre R', du mot binaire issu du prétraitement, c'est-à-dire du complément logique $\overline{X}_t = \left\{\overline{X}_t(i)\right\}$ de X_t. Les signaux de prédétermination de ce registre sont P_R et C_R, deux impulsions complémentaires décalées qui ont également servi à prédéterminer le registre R de l'ensemble de traitement.

Les opérations de dénombrement qui suivent et qui comprennent la formation de N_t et n_t se font par décalage des mots X_t et X_{t-1} stockés respectivement dans le registre R' et dans le registre IV, X_t étant restitué à partir de $\overline{X}_t$ en disposant un inverseur sur la sortie série du registre R'. Un train H'_d de 24 impulsions d'horloge H' synchronise le décalage.

Le nombre de transitions n_t est construit de la manière suivante :

- à chaque décalage, le bit $X_t(i)$ de X_t composé avec le bit $X_{t-1}(i)$ de X_{t-1} dans un opérateur OU exclusif (OU excl 1) détermine la valeur de $n_t(i)$;

- s'il y a transition, $n_t(i) = 1$ et le passage d'une impulsion H'_d vers l'entrée H VI du compteur VI est validé dans la porte III, provoquant l'incrémentation du compteur.

Ainsi, au terme du décalage, on aura formé dans celui-ci :

$$n_t = \sum_{i=1}^{24} n_t(i)$$

Le nombre N_t est généré dans le compteur V par validation d'impulsions H'_d sur l'entrée d'horloge H V de ce compteur, lorsque les bits $X_t(i)$ de X_t, appliqués successivement sur une des entrées de la porte II, valent 1.

Dans ce cas,

$$N_t = \sum_{i=1}^{24} X_t(i)$$

Notons que l'information X_t est sauvegardée dans le registre IV par lecture série de X_t durant le décalage de X_{t-1}. A l'instant $t+1$, cette information jouera le rôle de X_{t-1} à l'instant t.

Les registres V et VI contiennent, au temps t, les nombres respectifs N_{t-1} et n_{t-1} établis durant l'échantillonnage précédent dans les compteurs V et VI ; ces nombres sont transférés en parallèle dans les registres à la fin de la segmentation relative à cet échantillonnage.

A ce stade commencent les opérations arithmétiques d'addition et de division.

La somme $N_t + N_{t-1} = N^*_t$ est effectuée dans l'additionneur binaire I. La division par deux de N^*_t s'obtient en décalant ce nombre d'un bit vers les poids les plus faibles, c'est-à-dire en supprimant le bit de poids le plus faible. Le deuxième membre de l'inéquation $(x + 0,5\ N^*_t)$ est formé dans l'additionneur III, le seuil x étant composé manuellement, sur une molette d'affichage.

Le premier membre n^*_t correspond à la somme de n_t et de n_{t-1}, effectuée dans l'additionneur II.

Les deux membres sont ensuite comparés suivant le mode opératoire déjà explicité pour l'identificateur de maximum. Le résultat de la comparaison est donné par l'état de la variation binaire M_t :

- si $M_t = 1$, le régime est transitoire ($n^*_t > x + 0,5\ N^*_t$)
- si $M_t = 0$, le régime est soutenu ($n^*_t \leqslant x + 0,5\ N^*_t$)

La dernière opération consiste à détecter s'il y a eu ou non changement de régime par rapport au régime caractérisé lors de l'échantillonnage précédent. On compose à cet effet les variables M_t et M_{t-1} dans un opérateur OU exclusif (OU excl 2) pour former le bit de segmentation $S_{sg} = M_t \oplus M_{t-1}$ que l'on stocke dans la bascule B' du registre de sortie (l'information M_{t-1} a été rangée dans la bascule B au temps t-1).

Remarque :

Les deux membres dont on compare numériquement les valeurs ne sont pas en fait exactement construits comme on vient de le décrire. Nous nous sommes proposé, en effet, de doter le dispositif de segmentation d'un seuil variable afin de pouvoir le régler au mieux lors du traitement d'une séquence de parole ; or, tel qu'il est défini ici, ce seuil est ajustable à une unité près, ce qui correspond à une variation minimale trop élevée pour pouvoir approcher correctement le taux de segmentation optimal.

Nous avons partiellement remédié à cet inconvénient en doublant les grandeurs $0,5\ N^*_t$ et n^*_t qui interviennent dans l'inéquation, celle-ci s'écrivant alors

$$2\ n^*_t > x + N^*_t$$

Ainsi, à une variation unitaire du seuil dans cette inéquation, correspond une variation d'une demi-unité de celui-ci dans l'inéquation initiale. (Le nombre $2\ n^*_t$ se forme très simplement en ajoutant à n^*_t un bit fixé à 0, comme bit de poids le plus faible).

III.2.2. Synchronisme des opérations.

La segmentation, comme le traitement principal de reconnaissance, débute à chaque centième de seconde, par un transfert parallèle du mot binaire élaboré par le prétraitement, dans les registres R' et R. Les opérations dans les deux systèmes se déroulent ensuite jusqu'à ce que : pour le premier, l'information S_{sg} soit disponible et pour le second, le numéro du phonème reconnu soit rangé dans le registre de sortie.

Nous avons convenu d'affecter la valeur de S_{sg} établie au temps t , au numéro du phonème reconnu au temps t-1 (cf. paragraphe III.1). Cette information doit, par conséquent, être disponible avant que le contenu du registre de sortie ne soit modifié, donc avant que ne débute l'identification du maximum dans le traitement principal. La segmentation s'achève, pour cette raison, avant l'apparition de l'impulsion S_f de calibrage des fuites, le transfert de la nouvelle valeur de S_{sg} dans la bascule B' du registre III étant validé par cette même impulsion.

Une telle synchronisation exige un rythme de décalage plus élevé dans le registre R' que dans le registre R puisque le premier est totalement vidé de son contenu alors que le second ne l'est pas nécessairement (cf. paragraphe II.2.1.3.).

Par suite, la fréquence d'horloge H' choisie est deux fois plus grande que la fréquence d'horloge H définissant le rythme de décalage dans le registre R($f_{H'}$ = 12,8 kHz). Nous avons retenu ce rapport parce qu'à partir de H', il est très facile, en interposant une bascule, de générer H.

Les opérations qui succèdent au transfert de S_{sg} dans le registre de sortie sont des opérations de rangement ; chronologiquement (voir le diagramme des signaux de la planche 15), est ordonné :

- le transfert de M_t dans la bascule B par le niveau haut de l'impulsion r_8 de durée θ_8 ,

- le transfert, respectivement, de N_t et n_t dans les registres V et VI par deux signaux complémentaires décalés C'_R et P'_R.

Les signaux de prédétermination r_8 , C'_R et P'_R sont générés par les deux monostables 8 et 9 montés en cascade dont le premier est déclenché sur le front descendant de l'impulsion S_f.

Le tableau de la planche 16 résume les séquences d'opérations caractérisant d'une part, la reconnaissance de la parole, et d'autre part la segmentation phonémique de celle-ci durant le $t^{ième}$ échantillonnage ; la correspondance dans le temps entre ces deux séquences y est respectée.

III. 2. 3. Circuits intégrés utilisés.

Les composants utilisés dans cette réalisation sont sensiblement les mêmes que ceux décrits dans le deuxième chapitre ; ce sont tous des circuits intégrés de la famille TTL à faible ou moyenne intégration (respectivement SSI ou MSI).

Registres :

Les registres jouent ici le rôle de mémoires. On distingue :

- le registre R' de dimension 24 bits, similaire en tous points au registre R (cf. paragraphe II. 2. 1. 4).

- le registre IV (24 bits) formé de trois registres à décalage SN 7491 AN connectés en cascade ; le mot X_{t-1} est entré en série bit après bit, décalé dans le registre puis restitué en série sur la sortie ; le décalage s'opère sur le flanc montant des impulsions d'horloge.

- les registres V et VI qui mémorisent respectivement les nombres N_{t-1} et n_{t-1} , tous deux inférieurs ou égaux à 24 (registres à 5 bits SN 7496 N) et dont les entrées et les sorties se font en parallèle.

Compteurs :

Les compteurs V, VI et VII sont des compteurs binaires 5 bits (2 x SN 7493 N) réinitialisables à zéro.

Le compteur VII est associé à une porte de décodage qui verrouille la porte I et inhibe l'incrémentation de celui-ci après 24 impulsions d'horloge H'.

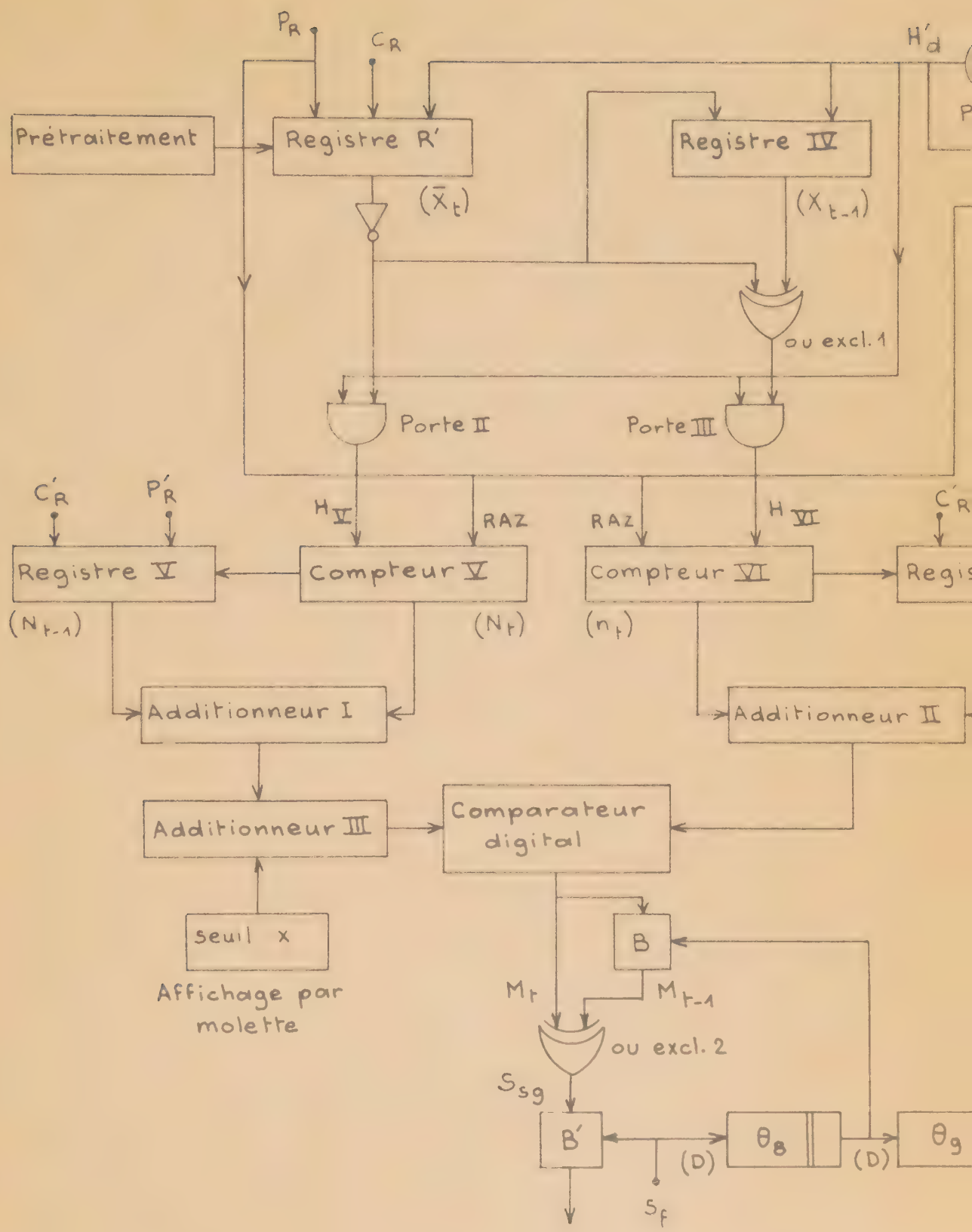
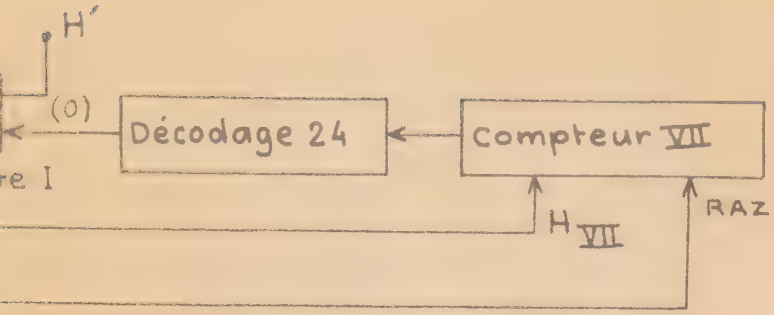
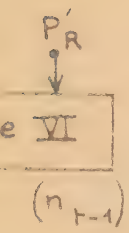


Schéma synoptique de





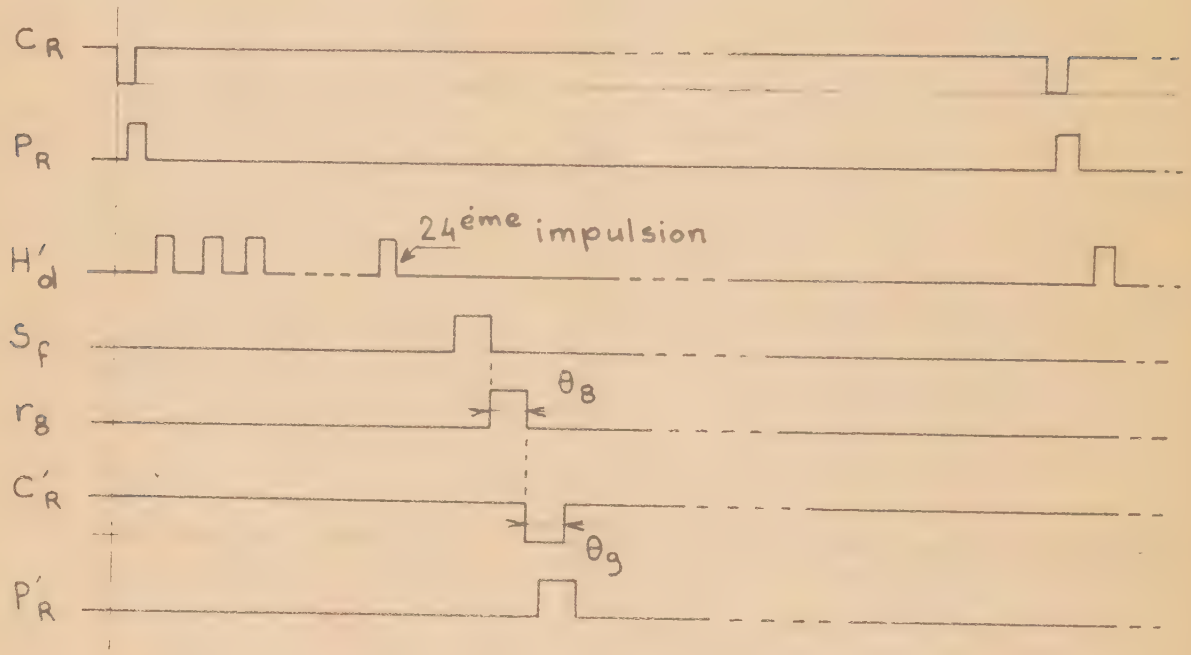
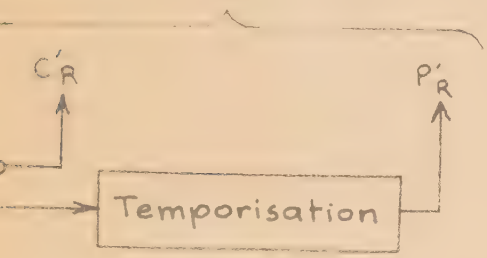


Diagramme des signaux

Prédétermination des registres V et VI



$$\begin{cases} \theta_8 = 2\,\mu s \\ \theta_9 = 2\,\mu s \end{cases}$$

segmentation

Planche 15

Planche 16.

Traitement principal	Transfert de $\overline{X}_t$ dans le registre R	Calibrages C_1 | C_2 | C_3	Fuites	Identification du phonème. Transfert de son numéro dans le registre de sortie.
Segmentation	Transfert de $\overline{X}_t$ dans le registre R'	Elaboration du bit de segmentation S_{sg}	Transfert de S_{sg} dans le registre de sortie. Saisie par ordinateur du mot stocké dans ce registre.	Rangement de M_t , N_t et n_t

temps →

Tableau comparatif des opérations effectuées lors du traitement principal et de la segmentation du signal de parole à l'instant t d'échantillonnage.

Additionneurs :

Les additionneurs I, II et III de cet ensemble sont chacun formés de deux additionneurs complets 4 bits type SN 7483 N branchés comme l'indique la figure 10 où $\sum = A + B$

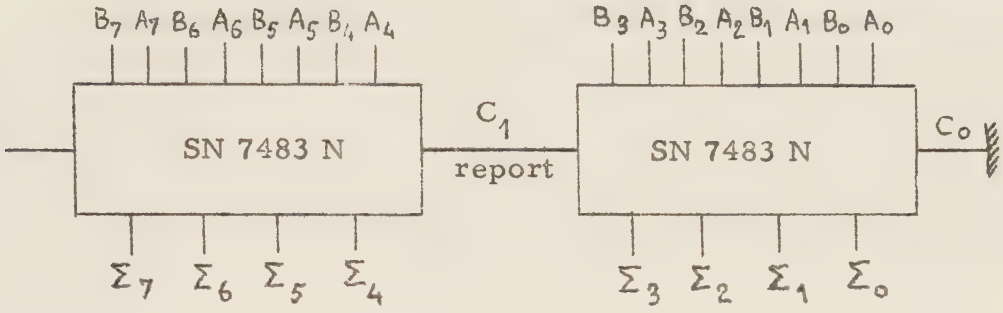


Figure 10.

Le report de l'addition s'effectue en série, les entrées A_i et B_i non utilisées étant au niveau logique zéro.

Les nombres binaires sommés en parallèle dans les additionneurs I et II ont pour dimension 5 bits ; le résultat de la sommation s'exprime sur 6 bits. Enfin, l'addition des grandeurs N_t et x, définies respectivement par 6 et 3 bits, se traduit sur les sorties de l'additionneur III par un nombre de 6 bits.

Comparateur digital :

Cette unité arithmétique exécute en parallèle le test $2n^*_t > N^*_t + x$ d'après le mode opératoire formulé au paragraphe II. 4. 4. La figure 11 montre comment les entrées du comparateur sont liées aux sorties des additionneurs II et III.

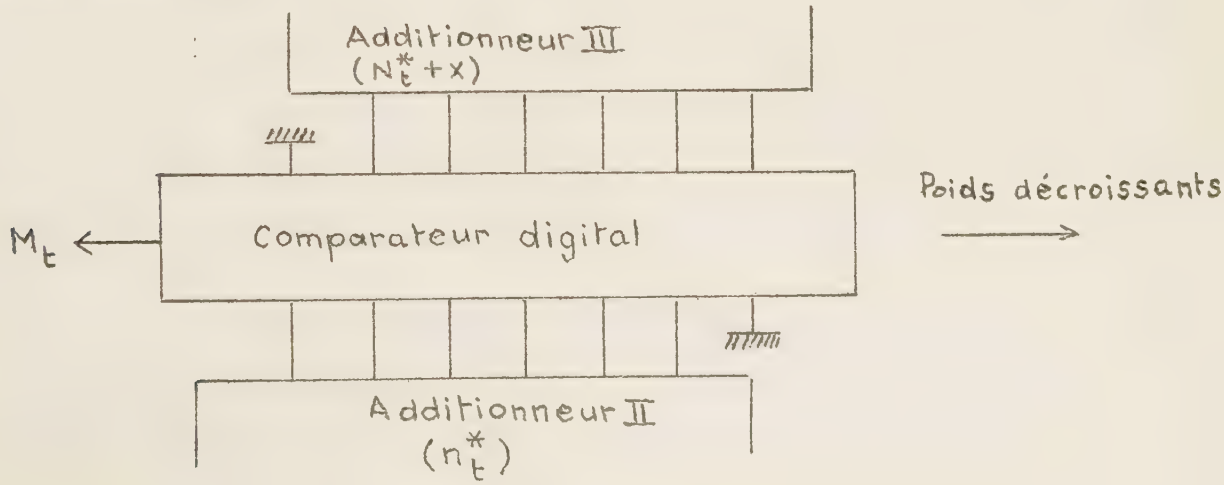


Figure 11.

CHAPITRE IV - EXPLOITATION DU SYSTEME DE RECONNAISSANCE ET DE SEGMENTATION - RESULTATS.

La configuration de l'ensemble réalisé implique l'utilisation d'un dispositif de visualisation du résultat stocké à chaque échantillonnage dans le régistre de sortie (régistre III , cf. planche 9). La fréquence d'apparition des informations dans ce régistre est telle qu'il est évidemment impossible, par affichage numérique, de disposer du résultat en temps réel.

Aussi la restitution des résultats est faite en temps différé sur l'imprimante d'un calculateur T 2000 Télémécanique. Les fonctions du calculateur se limitent ici à la saisie des informations, à leur stockage en mémoire centrale et en dernier lieu à leur impression sur la feuille du téléimprimeur. La durée de l'analyse d'une séquence de parole dépend donc de la place réservée en mémoire centrale pour le rangement des résultats.

Dans ce chapitre, nous expliquons d'abord brièvement comment s'opère la saisie des informations sur calculateur T 2000 , puis nous commentons les résultats obtenus en segmentation et en reconnaissance.

IV.1. Saisie des informations sur calculateur T 2000.

Le calculateur T 2000 est équipé d'un bac d'entrées-sorties permettant l'échange d'informations avec des périphériques (l'appareil de reconnaissance est ici considéré comme un périphérique). La liaison entre les deux ensembles est réalisée par l'intermédiaire d'un coupleur et d'un câble, le mode d'échange étant le mode parallèle.

D'une manière générale, l'entrée d'une information en mémoire centrale est synchronisée par plusieurs signaux de service émis, soit par le périphérique, soit , au niveau du coupleur, par le calculateur.

Elle se divise en trois phases :

- dans une première phase, l'appareil de reconnaissance envoie un signal d'interruption prioritaire (APPEL) qui signifie à l'ordinateur qu'une information demande son entrée ; l'exécution du programme en cours est alors stoppée ; le calculateur attend cette information ;

- dans une deuxième phase, la saisie de la donnée est effectuée lorsque le périphérique confirme par le signal $\overline{VAL}$, la disponibilité ou la validité de cette donnée ;

- enfin, le calculateur émet vers le périphérique une impulsion SERVI coïncidant avec la fin de l'acquisition ; cette impulsion peut servir à inhiber la validité de la donnée pour éviter une nouvelle acquisition de celle-ci.

Il est par conséquent nécessaire, afin de satisfaire aux contraintes liées au hardware du calculateur, d'élaborer les signaux APPEL et $\overline{VAL}$. Le premier n'est autre que le signal impulsionnel S_f commandant les fuites des intégrateurs ; il déclenche les interruptions prioritaires sur le flanc montant des impulsions. Le second est simplement formé sur la sortie d'une bascule D (SN7474 N) dont le montage est précisé figure 12.

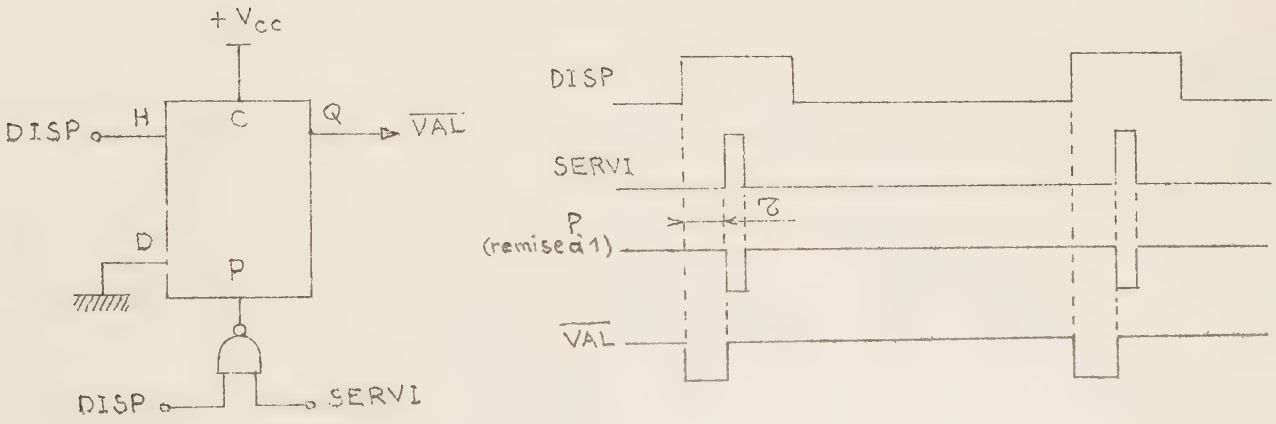


Figure 12.

Le signal $\overline{VAL}$ est construit à l'aide du signal DISP dont l'état haut matérialise la disponibilité d'une information sur les sorties du régistre III tant que ce niveau est maintenu. Ce signal s'identifie à l'impulsion r_8 et sert, sur son front de montée, à porter $\overline{VAL}$ dans l'état logique 0 qui, par convention, est l'état signalant à l'ordinateur la validité de l'information à acquérir.

Le programme proprement dit d'acquisition et le traitement des informations saisies est alors conçu comme l'indique l'organigramme de la figure 13.

Nous notons que ce traitement s'effectue uniquement quand il y a segmentation, le nombre des segmentations étant repéré par n. Quand n devient égal à N, nombre d'acquisitions effectives que l'on se propose de faire et qui est, rappelons-le, lié à la capacité en mémoire du calculateur, on masque les interruptions prioritaires afin de permettre l'impression des résultats sur le téléimprimeur.

IV.2. Résultats fournis par l'ensemble de segmentation.

Le calculateur a été un outil très efficace dans la vérification du bon fonctionnement de l'ensemble de segmentation et a permis en particulier le réglage du seuil x de façon optimale.

Cette vérification est basée sur le prélévement de la valeur du bit de segmentation S_{sg} et, simultanément, des données numériques fournies par le prétraitement. L'acquisition s' opère en deux temps sur deux coupleurs différents car, dans ce cas, le mot à saisir est formé de 25 bits (24 bits pour le prétraitement et S_{sg}), dimension excédant celle permise pour l'acquisition sur un seul coupleur. L'information est ensuite recadrée par programmation puis est imprimée dans le format représenté planche 17 : les astérisques figurent les bits portés au niveau logique 1, les blancs ceux portés au niveau 0, l'apparition de la segmentation est symbolisée par la lettre S. Nous reconnaissons une représentation équivalente au tableau de 1 et de 0 mentionné planche 1, le sens des fréquences

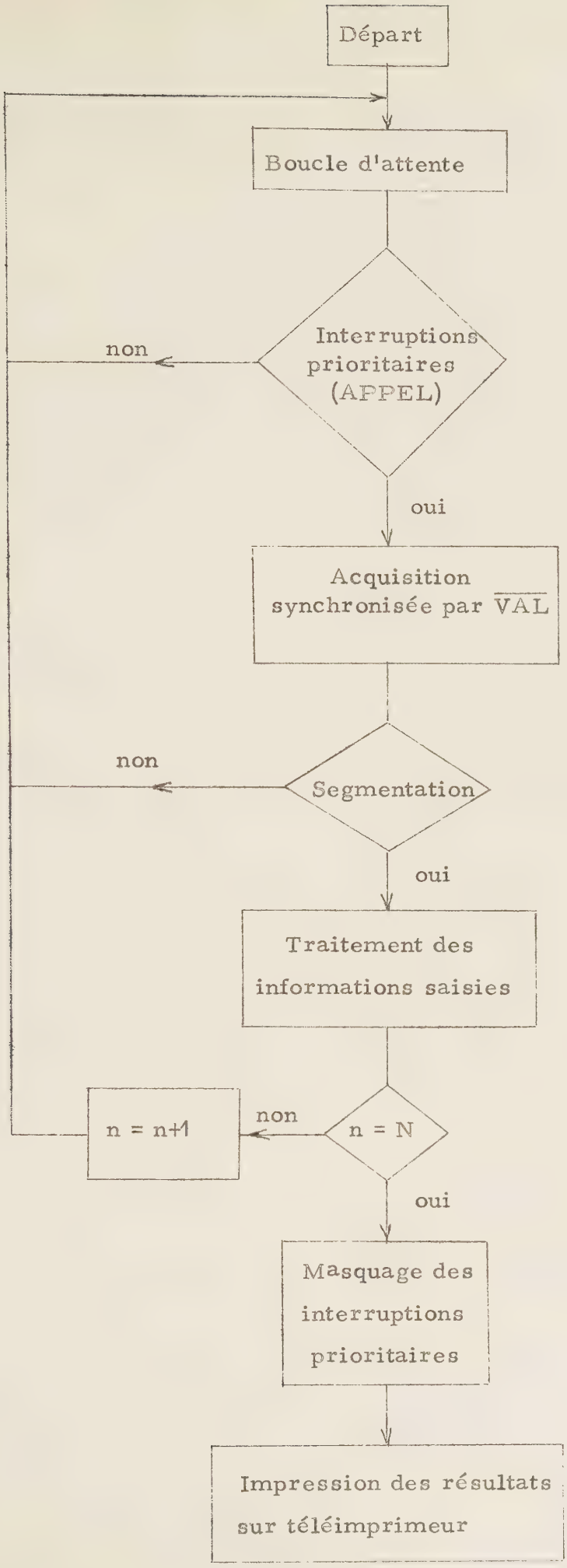

Départ
Boucle d'attente
Interruptions prioritaires (APPEL)
non
oui
Acquisition synchronisée par $\overline{VAL}$
Segmentation
non
oui
Traitement des informations saisies
n = N
non
n = n+1
oui
Masquage des interruptions prioritaires
Impression des résultats sur téléimprimeur

Figure 13.

croissantes étant ici, de la gauche vers la droite. Le spectre donné est celui du mot "intérêt" avec un seuil x fixé à 3.

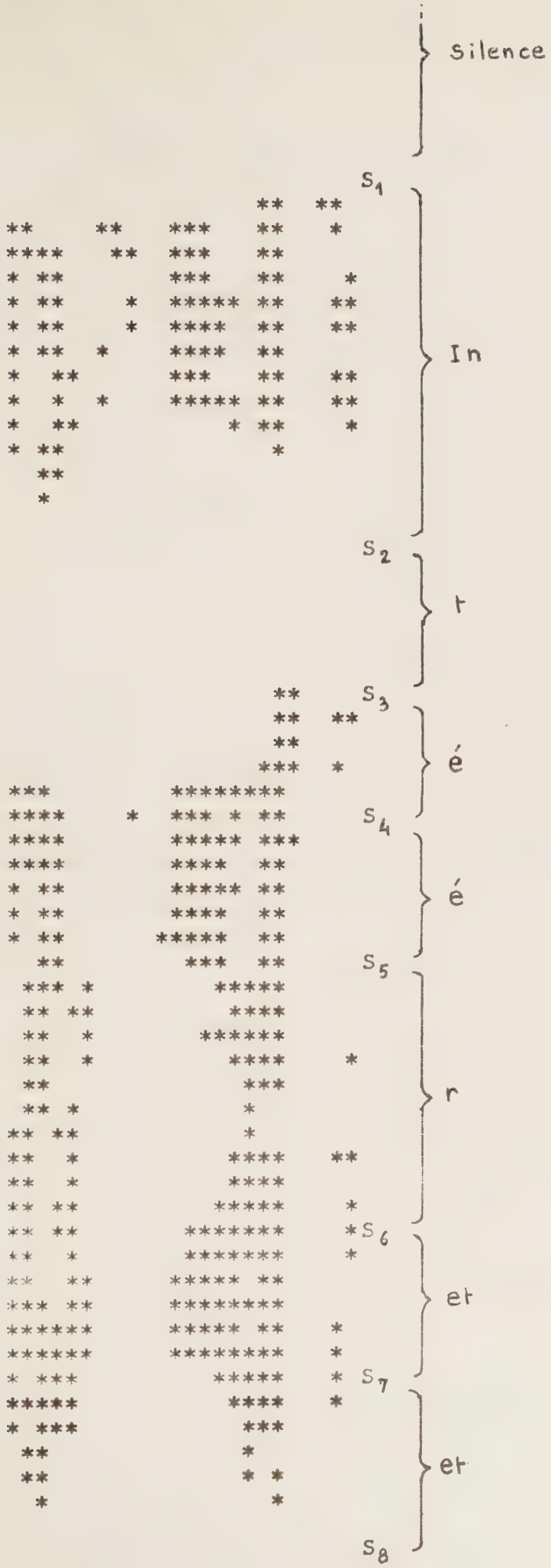
La segmentation en S_1 repère la transition entre le silence et le début du mot, puis une nouvelle transition est marquée en S_2 par l'apparition du silence précédant la prononciation du t ; ce silence se termine en S_3. La segmentation en S_4 signifie le début d'un régime soutenu (prononciation du é) jusqu'en S_5 où commence un régime transitoire (prononciation du r) ; de S_6 à S_7 le régime de parole peut s'assimiler à un soutenu (et) , S_8 coïncidant avec la fin du mot.

Le découpage du mot intérêt par l'ensemble de segmentation réalisé définit donc sept phonèmes si l'on exclut le silence initial. Ceci correspond, par rapport aux cinq phonèmes composant effectivement le mot à un taux de segmentation trop élevé. Mais cet état de faits concrétise bien le but recherché et ne présente pas d'inconvénients dans la mesure où le découpage caractérise bien l'apparence du spectre du mot.

On peut noter enfin que cette visualisation des données fournies par le prétraitement offre un avantage certain quand, par exemple avant tout apprentissage, il s'agit d'identifier chaque échantillon segmenté à un phonème. L'attribution des codes de phonèmes se faisant au clavier du téléimprimeur, on fera ensuite perforer sur bande l'ensemble de la séquence de parole traitée et sa segmentation.

IV. 3. Vérification du bon fonctionnement de l'ensemble de traitement - Résultats en reconnaissance.

Au niveau des résultats, la correspondance entre le dispositif de reconnaissance réalisé et celui simulé a également été vérifiée sur calculateur T 2000.



Segmentation du mot intéret

Planche 17

Dans ce but, deux programmes d'acquisition et de traitement en temps réel ont été mis au point :

- le premier se limite à la saisie du contenu du régistre III, à son décodage en décimal et à son impression sur la feuille du télétype ; c'est le phonème identifié par le hardware de reconnaissance lorsqu'il y a segmentation ;

- le deuxième simule, à partir du prétraitement, toutes les fonctions câblées de l'appareil mise à part la segmentation et aboutit à l'impression des résultats.

En fait, nous nous sommes rapidement aperçu que détecter la ligne de la matrice d'apprentissage présentant le potentiel le plus élevé, doncne s'intéresser qu'au contenu du régistre III , se révélait insuffisant (nous le montrerons par la suite). Aussi nous avons choisi d'introduire, dans les deux programmes mentionnés ci-dessus, un sous-programme de classement des six premières valeurs maximales rencontrées sur les sorties de la matrice.

Cette façon de procéder conduit :

- dans le premier cas, à faire la saisie de S_{sg} et des 8 bits du convertisseur analogique-digital trente fois par échantillonnage (on prendra alors pour signal DISP, le signal STATUS au lieu de r_8) ;

- dans le second cas, à acquérir S_{sg} et simultanément les 24 bits du prétraitement.

Le format d'impression des résultats est celui représenté planche 18, le phonème reconnu en première position étant placé à gauche. On donne, sur cette planche, deux tableaux figurant la reconnaissance des phonèmes composant le mot démanteler (ils sont théoriquement au nombre de huit) : le tableau A rassemble les résultats fournis par la

simulation, le tableau B, ceux fournis par l'appareil de reconnaissance. Le codage des phonèmes utilisé se réfère, pour une grande part, au code phonétique internationnal ; nous l'avons mentionné planche 19.

En comparant ces deux tableaux, il apparaît qu'un grand nombre de phonèmes classés en première et deuxième position dans le tableau A se retrouvent aux mêmes places dans le tableau B . Pourtant, chiffrer de manière exacte le degré de concordance des résultats s'avère difficile et de plus inutile. Nous nous contenterons donc de formuler, après un dénombrement étendu à une quarantaine de mots, que le phonème identifié en première position lors de la reconnaissance simulée est classé dans 75 % des cas à cette même position lors de la reconnaissance par le hardware, et qu'il se situe dans 90 % des cas, soit en première, soit en deuxième colonne de ces tableaux B .

Par conséquent, le système réalisé est conforme, avec une bonne précision, au système simulé, compte-tenu des erreurs qui peuvent s'introduire au niveau de l'affichage potentiométrique des tensions moyennes, de la matrice d'apprentissage dont les résistances ont été mesurées au pour-cent et du convertisseur analogique-digital dont la résolution se limite à 40 mV pour une plage de variation de 6 V ($\pm$ 3 V) lors des essais.

Ce travail s'arrête théoriquement au stade de la réalisation du système de reconnaissance et à la vérification de son fonctionnement ; c'est ce qui nous a préoccupé jusqu'alors. Il est néanmoins intéressant d'ajouter quelques remarques quant à la "qualité" des résultats obtenus en reconnaissance, qualité qui dépend essentiellement des concepts mis en oeuvre dans le dispositif.

Nous distinguerons deux groupes de résultats : ceux obtenus lors de la reconnaissance de mots déjà "appris" par le système, c'est-à-dire de mots dont les phonèmes ont servi de référence durant la phase d'apprentissage (cf. paragraphes I.2.1 et I.2.3) et ceux obtenus sur des mots non appris par le système.

TABLEAU A : par simulation.

22	13	18	14	4	29
22	29	15	20	4	6
15	29	22	20	4	9
29	22	15	9	4	6
9	29	6	15	22	30
9	22	4	15	29	26
1	11	13	25	10	18
5	14	7	22	13	18
14	7	22	29	23	28
9	5	14	7	18	8
1	7	13	18	14	19

TABLEAU B : par l'appareil de reconnaissance.

22	13	19	18	14	4
22	29	15	23	19	14
29	9	22	30	15	23
9	22	4	29	15	30
9	1	18	11	13	22
1	11	25	13	18	12
7	5	3	18	4	22
5	18	7	13	14	1
9	7	18	14	8	28
9	7	1	18	14	19
1	7	18	13	19	11

Lors d'un silence complet, les deux modes de reconnaissance établissent le classement suivant :

1 11 10 25 13 7

le silence est donc correctement identifié.

Code	Phonème	Mot(s) clé
1	Silence	
2	ɑ + a	plat
3	R	rue
4	l	lent
5	e	blé
6	s	sou
7	i	il
8	ɛ	lait
9	ə + œ	le, peu, heure
10	k	cou
11	t	ta
12	p	pas
13	d	dans
14	m	ma
15	ã	an
16	n	nous
17	U	ou
18	V	vie
19	Y + ɥ	nu, buit
20	õ	on
21	o + ɔ	eau, bol
22	ʒ	je
23	J	bailler
24	ɛ̃ + œ̃	lin, un
25	f	feu
26	b	bon
27	w	voir
28	g	gare
29	ʃ	chat
30	z	zéro

Code numérique associé aux phonèmes

Nous avons rassemblé, planche 20, les résultats de la reconnaissance des trois mots HALTERE , ALLURE et TERMINE . Les deux premiers appartiennent à la séquence d'apprentissage ; nous observons aisément que la reconnaissance phonémique de ces mots est beaucoup plus probante que celle relative au dernier mot :

- trois phonèmes des mots HALTERE et ALLURE sont reconnus en première position et on réalise un recouvrement pratiquement complet des phonémes en se limitant, sur ces exemples, aux deux premières colonnes (sauf pour ALLURE où le phonème 3 n'est pas présent) ;

- un seul phonème du mot TERMINE apparaît en première position et il faut classer jusqu'à la cinquième valeur maximale pour les recouvrir tous.

Néanmoins, dans les deux cas, certaines lignes sont totalement erronées. On en dénombre : une dans HALTERE , aucune dans ALLURE et trois dans TERMINE.

Au vu des quelques extraits de ces résultats obtenus, nous pouvons conclure qu'il semble impossible, à l'aide de notre dispositif, de reconnaître sur la première colonne un mot non appris et que cette tâche tient encore de la gageure si nous étendons la dimension du tableau jusqu'aux quatre ou cinquième colonnes. Cette constatation n'est pas surprenante car elle illustre bien le fait qu'un phonème est fortement influencé par le "contexte morphologique" dans lequel il est prononcé : si le système de reconnaissance n'a pas été sensibilisé à ce contexte, le phonème a d'autant moins de chance d'être reconnu. Les concepts mis en oeuvre dans le dispositif sont peut-être ici à mettre en cause car celui-ci, de part sa matrice d'apprentissage, est trop rigide pour s'adapter à des situations qui lui sont peu familières.

Reconnaissance de mots "appris"

HALTERE : transcription phonétique : 2 - 4 - 11 - 8 - 3

2	1	24	20	15	11
2	15	24	14	3	20
2	15	4	3	24	27
4	15	5	3	2	9
4	9	2	3	16	17
9	11	7	16	4	18
1	2	12	26	14	11
6	19	1	15	9	3
5	2	27	8	4	3
2	8	4	5	18	14
3	25	19	15	9	27

ALLURE : Transcription phonétique : 2 - 4 - 19 - 3

2	15	4	3	27	24
2	15	4	24	14	12
15	4	9	2	18	14
4	18	15	9	14	1
11	16	7	6	22	4
7	19	17	13	6	10
19	7	17	13	20	10
19	7	21	20	6	13

Reconnaissance de mots "non appris"

TERMINE : Transcription phonétique : 11 - 5 - 3 - 14 - 7 - 16 - 5

1	25	18	13	11	10
1	5	18	7	9	12
9	2	1	5	15	4
14	9	10	3	8	6
19	4	15	8	2	10
9	7	1	18	4	11
9	1	7	18	11	12
27	25	29	17	22	3
7	9	11	16	13	4
11	7	1	13	18	12
13	11	7	1	10	16
13	10	7	5	1	8
4	14	9	5	8	15

Les chances de succès sont beaucoup plus évidentes quand il s'agit de reconnaissance sur des mots appris. Les trois premiers phonèmes classés devraient permettre de discriminer un mot parmi tous ceux composant la séquence d'apprentissage. Sous cet aspect, la reconnaissance consisterait donc en une recherche de mots dans un dictionnaire, les mots de la séquence d'apprentissage étant ceux du dictionnaire.

* * * * * * * *

CONCLUSION

Les deux parties du processus de reconnaissance donnent des résultats conformes à ceux que l'on espérait.

L'ensemble de segmentation détermine, avec précision, et de manière simple, les zones de régime soutenu et de régime transitoire dans les spectres discrets fournis par le prétraitement numérique. Ce hardware sera utilisé dans les recherches ultérieures.

L'ensemble de reconnaissance proprement dit réalise, avec une bonne concordance, les fonctions simulées que l'on s'était proposé de câbler, ceci compte-tenu de la dégradation des informations inhérentes aux parties analogiques du système (affichage, matrice d'apprentissage, convertisseur).

Nous avons pourtant fait quelques restrictions quant à la possibilité d'utilisation des résultats obtenus. Il apparaît que le processus est peu efficace lors de la reconnaissance de mots non appris, et que, même lorsque celle-ci s'effectue sur des mots appris, le phonème souhaité n'est reconnu que dans 50 % des cas en première position.

L'exploitation du système doit pourtant être possible pour la reconnaissance de mots d'un dictionnaire limité, la séquence des mots de ce dictionnaire devant impérativement constituer la séquence d'apprentissage. Des essais sont en cours afin de mettre au point, à partir des trois premiers phonèmes classés, un algorithme de recherche du mot prononcé. Néanmoins, il subsiste que l'utilisation d'un système à apprentissage définitivement figé prêtera à nouveau à défaillances dès lors que le nombre de mots du dictionnaire augmentera : des recherches se poursuivent en vue "d'assouplir" les fonctions de discrimination du processus.

Enfin, nous tenons tout de même à noter que l'achèvement de cette réalisation a permis à notre groupe de recherches de se familiariser avec le traitement d'informations en temps réel, traitement qui nous était interdit ces dernières années encore et qui, de ce fait, retardait l'évolution des recherches.

* * * * *

BIBLIOGRAPHIE.

[1] N. NILSSON , Learning Machines, New-York : Mc-Graw-Hill , 1965.

[2] G. SEBESTYEN , Decision-Making Processes in Pattern recognition, ACM Monograph, New-York : Macmillan , 1962.

[3] M. J. DUGRAVOT , Prétraitement numérique du signal vocal dans le domaine spectral, thèse 3ème cycle , 1969.

[4] M. J. DUGRAVOT , M. LAMOTTE , Prétraitement numérique du signal vocal, l'Onde Electrique , n° 514 , p. 69-71 , 1970.

[5] Lernende Automaten, Bericht über die Fachtagung der Nachrichtentechnischen Gesellschaft im VDE (NTG) , Fachausschuss 6 "Informationsverarbeitung" in Karlsruhe am 13. und 14. April 1961.

[6] G. NAGY , State of the Art in Pattern Recognition, Procedings of the IEEE , vol. 56 , n° 5 , May 1968.

[7] A. FRUHLING , M. LAMOTTE , M. J. VIGNERON , J. BREMONT , J. P. HATON , noter au VI ème Congrès International de cybernétique , Décembre 1970.

[8] M. J. VIGNERON , M. LAMOTTE , J. BREMONT , J. P. HATON , Recherches actuelles sur l'extraction de caractéristiques et la reconnaissance de la voix parlée , Automatisme Décembre 1970 , p. 646-649.

[9] B. ROGER , M.F. TOUTAIN DE BUSSY , Applications des amplificateurs opérationnels , Notice Imex - France.

[10] J.C. MARCHAIS , L'amplificateur opérationnel et ses applications , Masson et Cie 1971 , chap. III , p. 33-41.

NOM DE L'ETUDIANT : HORNUNG Robert

Nature de la Thèse : DOCTEUR-INGENIEUR.

Vu, Approuvé

et permis d'imprimer.

NANCY, le 31 JAN. 1973

Le Président du Conseil de l'Université

de NANCY I,

J.R. HELLUY